Investigating Equity and Social Justice within Programs: A Framework of Evaluative Criteria

Rebecca M. Teasdale
University of Illinois Chicago

Cherie M. Avent
University of Illinois Urbana-Champaign

Background: Evaluators have a professional and ethical responsibility to contribute to the "advancement of an equitable and just society" (American Evaluation Association [AEA], 2018b). A rich body of scholarship provides guidance about how evaluators can do so through culturally responsive, social justice–oriented and decolonizing evaluation approaches, as well as culturally responsive methods, attending to power and privilege in program and evaluation contexts, and partnering with communities. In this article, we provide guidance for examining how the program being evaluated attends to issues of equity and social justice.

Purpose: We present a framework for investigating equity and social justice within programs through the criteria that evaluations pursue. The framework is offered as a map of possibilities and a thinking tool to help evaluators surface, examine, and negotiate varying values and design evaluative lines of inquiry to address them.

Setting: We write as evaluation faculty in colleges of education. We approached the framework through the lens of our shared commitment to advancing equity and social justice in evaluation practice. We aimed to bring together our differing areas of expertise and lived experiences to develop a resource to support evaluators in advancing equity and social justice.

Intervention: Not applicable.

Research Design: Not applicable.

Data Collection and Analysis: Our work builds on a previously published model of criteria domains. Each domain reflects a broad category of program characteristics or results that can be deemed important or desirable for a given program and context. Here, we refine and expand the model by applying an equity and social justice lens to 11 different criteria domains. We draw on the social justice–oriented evaluation literature for guidance and examples about how evaluators can advance equity and social justice within each domain and associated lines of inquiry.

Findings: The framework outlines 11 criteria domains in which a program's contribution to equity and social justice might be examined. We describe each domain and apply it to an example evaluation to illustrate. We conclude by discussing the use of the framework to advance equity and social justice through evaluation practice, education, and scholarship.

Keywords: equity; social justice; criteria; values

Evaluators have a professional and ethical responsibility to contribute to the "advancement of an equitable and just society" (American Evaluation Association [AEA], 2018b). This responsibility is rooted in evaluation's role in helping society deepen understanding of social issues and possible approaches to address them (Cronbach & Associates, 1980; Greene, 2006; Thomas & Campbell, 2021). Ultimately, evaluation can be used to inform action to address complex problems and improve social conditions (M. Hall, 2020; Hopson & Cram, 2018; Rossi et al., 2019).

Many social issues and conditions are produced and upheld through the destructive forces and consequences of interlocking systems of oppression. These include colonization and colonialism (Chilisa et al., 2016; Waapalaneexkweew, 2018), racism (Hall, 2018; Thomas et al., 2018), sexism (Hood & Cassaro, 2002; Sielbeck-Bowen et al., 2002), ableism and marginalization of people with disabilities (Jacobson et al., 2012; Mertens, 1999), socioeconomic inequities and class oppression (AEA, n.d.), and homophobia, transphobia, and other oppression based on sexual orientation and gender identity (Miller, 2018; Phillips et al., 2023). Evaluators have highlighted the critical role evaluation can play in dismantling systems of oppression and fostering positive change and healing (AEA, n.d.; Canadian Evaluation Society, n.d.; McBride et al., 2020; Neubauer et al., 2020). Specifically, evaluation can be used to provide evidence about programs and contexts, foster learning, and improve decision-making to promote "healthy, just, and equitable communities" (Thomas & Campbell, 2021, p. 6).

A rich body of scholarship provides guidance about how evaluators can advance equity and social justice in their practice. This includes literature that advances culturally responsive (e.g., Hood et al., 2015; Manswell-Butty, 2004; Thomas, 2004), social justice–oriented (e.g., Mertens & Wilson, 2019), and decolonizing (e.g., Chilisa & Bowman, 2023; Cram, 2016) evaluation approaches and frameworks—and outlines the shared commitments that underpin these approaches (e.g., Jordan & Hall, 2023; Thomas & Campbell, 2021). Further, the literature addresses culturally responsive evaluation methods (e.g., Bowen & Tillman, 2015; J. Hall, 2020a), attending to power and privilege in program and evaluation contexts (e.g., J. Hall, 2020b; Thomas & Campbell, 2021), and partnering with communities who are involved with and affected by the program and focal issues (e.g., Bledsoe, 2021; Madison, 1992; McBride et al., 2020). Evaluators can also focus their inquiry to examine how the program, policy, or other initiative being studied (i.e., the evaluand) itself advances equity and social justice or reinforces inequity and injustice (Giacomini & Hurley, 2008; Rogers, 2016; Teasdale, 2021). To do this, evaluators can work with those involved and affected to identify the evaluand characteristics or results that are important or necessary for the evaluand to advance greater equity or social justice within the focal community (Bledsoe, 2021; Teasdale et al., 2024; Thomas & Campbell, 2021). These desired characteristics or results can serve as evaluative criteria that define a "high-quality" or "impactful" evaluand (Teasdale, Pitts, et al., 2023). Criteria, in turn, guide the evaluator's line of inquiry by shaping the evaluation questions, performance indicators, and methods they pursue and the conclusions they reach (Patton, 2021; Teasdale, 2022a).

To date, there is limited literature about the specific criteria and lines of inquiry evaluators might investigate when examining issues of equity and social justice within the evaluands they study. Thus, there have been calls for scholarship to deepen the knowledge base about equity- and social justice-focused criteria and lines of inquiry (Gates, Williamson, et al., 2022; Teasdale, 2022a; Teasdale, Strasser, et al., 2023). In response, we present a framework in this paper to support and guide evaluators in defining equity- and social justice–focused criteria and using those criteria to shape evaluation questions, performance indicators, and evaluative conclusions. We begin by discussing our positionality and the context for this article. Next, we review the literature on evaluative criteria and explain how we developed the framework. We then outline the framework by describing 11 domains in which a program's contribution to equity and social justice might be examined and applying each domain to an example evaluation. We conclude by discussing the use of the framework in evaluation practice, education, and scholarship.

Context and Author Positionality

We are evaluation faculty in colleges of education. We began our collaboration in the fall of 2020 in the context of the COVID-19 pandemic and uprisings for racial justice across the United States and around the world. At that time, many social and educational programs and funders were grappling with their roles in systems of oppression and how they could advance equity and social justice. We understood that evaluation can play a critical role in advancing equity and social justice for programs. Thus, we embarked on our research partnership to

elevate and learn from social justice–oriented evaluators—specifically, those who had long worked to advance racial equity and justice—and develop resources to support evaluation practice.

Rebecca Teasdale identifies as a white, queer, abled, cisgender woman. Her interest in the topic stemmed from her efforts to advance equity and social justice in her evaluation practice and her research on evaluative criteria. Through her scholarship, she seeks to make values and definitions of quality more explicit in evaluation practice and support evaluators in systematically identifying, negotiating, and applying criteria. Cherie Avent identifies as a Black, abled, cisgender woman, spouse, and mother. In her scholarship she investigates the intersections of social justice–oriented evaluation and influences on evaluator practices. Her interest in the subject is rooted in a commitment to centering culture and social justice in her evaluation research and practice. We approached the framework reported here through the lens of our shared commitment to advancing equity and social justice in evaluation. We aimed to bring together our differing areas of expertise and lived experiences to develop resources to support the field in fulfilling our collective responsibility to the "advancement of an equitable and just society" (AEA, 2018b).

Evaluative Criteria

We conceptualize evaluation as a practice of examining the quality (i.e., value, merit, significance) of an evaluand to deepen understanding, identify improvements, or fulfill accountability requirements (Schwandt & Gates, 2021). This requires understanding what "quality" means for the specific evaluand, context, and constituents. [1] In evaluation theory, these definitions of quality are referred to as evaluative criteria (Davidson, 2005a). Criteria represent people's visions of a high-quality evaluand by describing the characteristics or results that are deemed important or desirable (Davidson, 2005b; House, 2015; Mark et al., 2000; Schwandt, 2015). Thus, criteria represent values about what matters most (Gates et al., 2024).

Criteria are often associated with Scriven's (2007, 2012) logic of evaluation. In this framing, evaluators establish criteria of merit, set standards for each criterion, and compare the evaluand's performance to the standards (Fournier, 1995). In

this view, criteria are understood as evaluand-specific indicators of quality. Ozeki et al. (2019) note that "synonyms sometimes used by evaluators to describe criteria include indicators, measures, or variables" (p. 1).

Evaluative criteria also function at a broader level. In interviews with evaluators, Stake et al. (1997) found that:

> evaluators spoke of criteria but not so much as succinct and critical particulars—as Scriven prescribed—but more as broad classes in which data can be placed and easily referred to [...] Seldom are the criteria seen as direct criteria of merit but rather information categories from which interpretations of merit are made. (p. 92)

That is, criteria are understood as broad dimensions of quality that evaluators use to represent and organize different perspectives on quality. Addressing this level, criteria are also referred to as dimensions of success, quality domains, and criteria domains (Teasdale, Moore, et al., in press).

We use an inverted triangle (Figure 1) to represent the multiple levels on which evaluative criteria operate (Sadler, 1985; Teasdale, 2022b). At the wide top, criteria function as broad quality categories. These dimensions of quality help evaluators distill and make explicit the essence of what matters most for the context and constituents—and then shape lines of inquiry to address the relevant evaluand characteristics or results. At the narrow bottom of the triangle, criteria function as evaluand-specific indicators of quality. These concrete definitions of quality are derived by operationalizing the broad categories at the top of the triangle and then used to guide data collection and synthesis of evidence to reach conclusions.

Evaluation professional organizations in North America have called on evaluators to engage with values transparently (AEA, 2018a, 2018b; Yarbrough et al., 2011). Thus, for both levels, evaluators are urged to select and articulate explicit criteria (Davidson, 2005b; Greene et al., 2011; Roorda & Gullickson, 2019; Schwandt, 2015; Teasdale, Pitts, et al., 2023). Yet, in practice, evaluative criteria—and the values they represent—often remain unstated and implicit (Greene, 2012; Gullickson, 2020; Gullickson & Hannum, 2019; Teasdale, Strasser, et al., 2023). To support explicit

---

[1] In this article we use the term "constituents" to refer to community members, program participants, staff, program leaders, evaluation commissioners, and

evaluators who are involved with and affected by a specific evaluand and evaluation.

attention to criteria, scholars have developed frameworks that outline quality categories for evaluators to consider (e.g., Organisation for Economic Co-operation and Development [OECD], 2019; Teasdale, 2021) and describe ethical perspectives for guiding criteria selection (Roorda & Gullickson, 2019). Rubrics are also advanced as a means for explicating evaluand-specific criteria, defining "good" performance, and guiding synthesis of evidence and criteria to reach conclusions (Davidson, 2005b; Dickinson & Adams, 2017; Martens, 2018; King et al., 2013). (See Montrosse-Moorhead (2022) for additional methods of synthesis across criteria.)
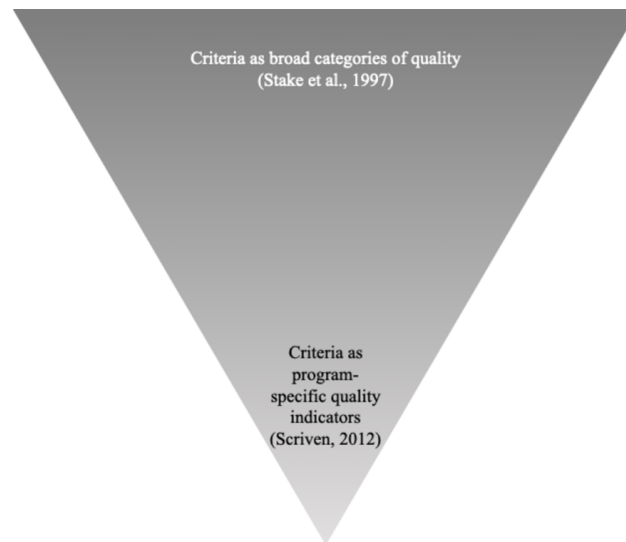
Model of Criteria Domains

Our work builds on Teasdale's (2021; Teasdale, Pitts, et al., 2023) model of evaluative criteria, which describes two aspects: domain and source. Domains refer to the focus or substance of a criterion by describing high-level categories of evaluand characteristics or results that can define quality. Sources describe the individual, group, or document from which criteria can be drawn. In this article, we focus primarily on criteria domains.

The domains in the model were identified through synthesis of existing criteria guidelines,

followed by empirical research. To begin, Teasdale (2021) synthesized three bodies of guidelines: (1) frameworks from the Organisation for Economic Co-operation and Development (OECD) and its Development Assistance Committee (DAC; OECD, 1991, 1999, 2002, 2018),[2] (2) guidelines from the Canadian and U.S. federal governments (Centre for Excellence for Evaluation, 2015; Dumaine, 2012; Shipman, 2012; Treasury Board of Canada Secretariat, 2012, 2016; United States General Accounting Office, 1988), and (3) guidance from evaluation scholars (Armytage, 2011; Chianca, 2008; Davidson, 2005b; Greene et al., 2011; Hurteau et al., 2009; Schwandt, 2015; Scriven, 2000). Next, a series of empirical studies were conducted to refine the list of domains. These included analysis of evaluation reports (Teasdale, 2021, 2022a) and peer-reviewed evaluation studies (Teasdale, Strasser, et al., 2023).

The resulting list of domains (Table 1) describes high-level categories of quality with relevance for a range of evaluands. The domains are defined broadly so they can be adapted and operationalized to align with specific evaluands, contexts, and constituents. The list is intended not to be prescriptive but to serve as a map of possibilities (Teasdale, 2021).

Figure 1. Evaluative Criteria at Multiple Levels

Criteria as broad categories of quality
(Stake et al., 1997)

Criteria as program-specific quality indicators
(Scriven, 2012)

---

[2] The OECD/DAC guidelines were revised in 2019 (OECD, 2019) to incorporate previous guidance for evaluations of humanitarian assistance (OECD, 1999).

The humanitarian assistance criteria had already been included in the synthesis described above.

Table 1. Criteria Domains

Domain	Essential question	Description
Alignment	How well does the evaluand complement or fit with other programs, initiatives, policies, etc.?	Evaluand is consistent and coordinated with larger or related initiatives, programs, or policies.
Design/ Implementation	Is the evaluand well-conceived and well-executed?	Conceptualization and execution of evaluand are consistent with relevant theory, best practices, or requirements. Evaluand is conducted as planned. Evaluand fits the context.
Equity	Does the evaluand advance equity?	Opportunities, experiences, benefits, and results are fair and just. Evaluand attends to historical or structural inequities. Evaluand prioritizes minoritized populations.
Experience/ Relevance	Do people have a positive experience with or positive attitude about the evaluand?	Evaluand activities are delivered in a way that is respectful, rewarding, or enjoyable. Participants or staff have a positive opinion or attitude about evaluand or its activities. Participants or staff are willing to adopt evaluand or choose to incorporate it into their ongoing activities.
Outcomes/ Impact	Did the evaluand work? Did it yield desired benefits?	Evaluand achieves its desired short-term or long-term outcomes, impact, effects, or results.
Reach/ Access	Did the evaluand serve and retain the intended participants?	Evaluand serves and retains intended participants. Intended participants are able to engage with the evaluand without undue barriers.
Relevance	Was the evaluand well-matched to participants' or community needs and circumstances?	Evaluand aims and activities are consistent with the needs, requirements, culture, interests, or circumstances of its intended participants.
Replicability	Could the evaluand work elsewhere?	Evaluand components, activities, or its underlying model or principles can be duplicated or adapted to another context.
Resource Use / Resource Allocation	Was the evaluand cost-effective? Was the evaluand adequately funded?	Evaluand funding, personnel, and materials are used economically. Funding, personnel, and materials are sufficient for implementation. Evaluand yields an appropriate level of benefit in relation to the funds, personnel, and materials required.
Sustainability	Are the benefits long-lasting?	Evaluand has (or is likely to have) long-term benefits through lasting outcomes or continuation of activities.
Unintended Effects	Did the evaluand have side effects?	Evaluand is associated with absent or minimal negative consequences or with unplanned positive consequences.

Adapted from Teasdale, 2021; Teasdale, Pitts, et al., 2023; and Teasdale, Strasser et al., 2023.

Equity as an Evaluative Criterion

The model presented in Table 1 includes Equity as one of 11 criteria domains. In this article, we refine and expand the Equity domain to assist evaluators in defining and applying equity and social justice as an evaluative criterion in their practices.

In the model (Teasdale, 2021; Teasdale, Strasser, et al., 2023), the Equity domain addresses whether and how (a) opportunities, experiences, and outcomes are fair and just; (b) the evaluand

addresses historical or structural inequities; and (c) the evaluand centers minoritized populations. [3] Conceptualizations of equity are highly contextual (Avent, 2025, Avent et al., 2023; Boyce et al., 2023). Therefore, rather than imposing a particular definition of equity, the domain directs attention to the issues of equity that are most salient for the evaluand, context, and constituents.

Other scholars and frameworks have also addressed equity as an evaluative criterion. For example, Kirkhart (2016) explains:

> ...equity may be used in the evaluation as a criteria of merit of the evaluand. This may apply to program process (e.g., equity of access to or experience with program services) or program outcome (e.g., equitable impacts of services across diverse consumers). To what extent has inequity been addressed and reduced by this evaluand (or have inequities been maintained or even increased)? (p. 116)

Boyce and Smith (2021) define the criterion of equity as "parity in program access, participation, and accomplishment for all program participants, especially those least well-served in the context (Greene, Boyce, & Ahn, 2011)" (p. 2). Gullickson and Hannum (2019) argue that using equity as a criterion allows evaluators to check for alignment between the values a program espouses, the values enacted in its practices, and the values realized in its outcomes.

Rogers (2016) reports that much of the focus on equity as a criterion centers on measuring unequal program results. However, in an empirical study, Gates and colleagues found that evaluators defined the criterion of equity in terms of program reach and access, differential experiences and outcomes, and root causes of inequities (Gates, Madres, et al., 2022). Further, Giacomini and Hurley (2008) report that equity is widely embraced as a criterion for evaluating health programs and policies by addressing resource distribution and decision-making. Boyce and colleagues (2023) found that evaluators working in science, education, technology, and mathematics (STEM) education contexts more often investigated the criterion of diversity, which addresses demographic variety among program participants and leaders, rather than equity. Boyce and colleagues (2023) also reported it was unclear whether STEM education evaluators investigated the criterion of inclusion, defined as the extent to which "participants are and

feel welcomed, embraced, included, and valued as learners" (p. 68). They urged greater attention to the criteria of equity and inclusion in order to assess "programmatic attention toward diversity, equity, and inclusivity" (p. 69).

In contrast to frameworks that include equity as an evaluative criterion, the revised OECD/DAC framework for international development (OECD, 2019) does not list equity as a stand-alone criterion. Instead, evaluators are encouraged to conduct equity-focused analysis within investigations of the relevance and effectiveness criteria. For example, when evaluating effectiveness, evaluators are urged to look beyond overall outcomes to examine differential outcomes and outcomes for marginalized groups. This mirrors empirical findings from Teasdale (2022a; Teasdale, Strasser, et al., 2023) that suggest the Equity domain may function as a lens that cuts across other domains (e.g., equity of outcomes, equity of experiences).

As part of our larger collaboration, we investigated criteria in evaluations of programs in which racial equity and justice played a role (Teasdale et al., 2024). Our findings confirmed that evaluators do examine the criterion of Equity within and across multiple criteria domains (e.g., equity of access, equitable outcomes) and expanded the scope to include both Equity and Social Justice, reflecting the language evaluators used. Our findings began to explicate an emerging framework of how evaluators can apply the Equity / Social Justice criterion as a cross-cutting lens.

In this article, we present a fully developed framework. We aim to address gaps in evaluator training related to criteria (Teasdale, Pitts, et al., 2023), which may contribute to the limitations in understanding and use of criteria documented by Ozeki et al. (2019). Specifically, we offer a resource to support evaluators in negotiating differing perspectives on equity and social justice (Avent et al., 2023), navigating preset lines of inquiry that may not adequately address equity and social justice (Gates, Williamson, et al., 2022; Teasdale et al., 2024), and identifying appropriate equity- and justice-focused indicators and metrics (Teasdale et al., 2024).

Framework of Criteria for Investigating Equity / Social Justice Within Programs

Table 2 presents the framework for applying the criterion of Equity / Social Justice as a lens that cuts

---

[3] We use the term "dominant" to refer to social groups afforded more power, privilege, and resources than others (AMA, 2021; J. Hall, 2020b). We use the term "minoritized" to refer to groups afforded less power, privilege, and resources through active processes of oppression (AMA, 2021; Benitez, 2010).

across other criteria domains. In this section, we describe our development process. We then summarize each domain in the framework and apply it to an example evaluation.

Table 2. Framework for Applying the Criterion of Equity / Social Justice as a Lens Across Other Criteria Domains

Criteria domain	Equity / Social Justice domain applied as a cross-cutting lens	
	Essential questions	Guidance and examples from the evaluation literature
Alignment	• In what ways is the evaluand part of a coordinated, longer-term strategy to advance equity and social justice? • How does it build on or advance other equity-focused work? • How does it align with broader priorities regionally, nationally, or globally?	What is the value of the intervention in relation to […] equity-focused national priorities and national and international partners' equity-focused policies and global references such as human rights? (Bamberger & Segone, 2011, p. 35) Did the project build upon or incorporate previous inclusion efforts in the community, and if so, how? (Elam & Walker, 2021 p. 219)
Design/ Implementation	• How does the evaluand take the root causes of inequities into account? • In what ways does the evaluand attend to both immediate needs and larger systems and structures? • In what ways does the underlying program theory reflect deficit perspectives regarding minoritized populations? • How does the evaluand employ inclusive, equitable, culturally sustaining, or culturally relevant practices? • How do the design and activities reflect the values and perspectives of minoritized populations? • In what ways were context and culture taken into account in evaluand design? • How were equity goals and processes incorporated into the planning of the evaluand and/or reflected in the program theory? • How were culture, language, or race considered in the development of the evaluand?	Too often program theory is based on perceived pathologies of the communities and families rather than the "root causes" of the conditions of social and economic impoverishment […] Evaluators must examine the systems and institutions that sustain an inequitable distribution of resources and opportunities. (Thomas et al., 2018, p. 523) What are the power dynamics operating within the project or between the project and the surrounding community? To what extent was the project implemented fairly, ethically, culturally appropriately, and in a way consistent with legal and professional standards? (Thomas & Campbell, 2021, p. 317) How and in what ways are project leadership attending to inclusion and cultural issues across components? (Boyce & Smith, 2021) More specific evaluation questions can be asked about the extent to which processes address [equity and gender equality], and about the results: To what extent were gender equality goals and processes incorporated into the planning of the intervention? To what extent did the intervention promote equal access by men and women to the benefits of the activity to resources, services and skills? (Rogers, 2016, p. 207) How did logistics and execution of activities take into consideration the community and cultural

	• To what extent is this strategy likely to be effective in reducing inequities? • What is the quality of services for the minoritized populations? • Are there disparities in services received between members of minoritized and dominant groups?	context of the priority population? (Elam & Walker, 2021, p. 23) How did the grantee work to avoid disparities in the services received by different racial and ethnic groups? (Elam & Walker, 2021, p. 27) [Evaluators] assessed the extent to which program staff shared social identities with focal communities […] Evaluators also examined the extent to which program leaders and staff demonstrated an understanding of social justice and the capacity to do equity-focused work, the extent to which programs engaged (or did not engage) communities of color, and how programs used community input in program design. (Teasdale et al., 2024, pp. 15–16)
Experience	• How do members of minoritized groups experience the evaluand? • Are there disparities in satisfaction between members of minoritized and dominant groups? • Are there disparities in participants' experiences or in how they are treated? • Are there disparities in willingness to engage with the evaluand or the acceptability of the evaluand between members of minoritized and dominant groups?	Third, equity may be used in the evaluation as a criterion of merit of the evaluand. This may apply to program process (e.g., equity of access or experience with program services). (Kirkhart, 2016, p. 116) How do different groups of participants experience the project activities? (Thomas & Campbell, 2021, p. 317) What is the project culture and climate? What are participant experiences and sense of belonging? Are there differences in experience across groups? (Boyce & Smith, 2021) Evaluators investigated whether participants had positive experiences in a program, were satisfied, and felt safe—and disaggregated data to identify disparities. (Teasdale et al., 2024, p. 16) These early African American evaluators collected qualitative and quantitative data to reveal differences in the educational experiences of African Americans and Caucasian Americans. (Hood, 2001, p. 36)
Openness	• To what extent are evaluand staff and leaders open to designing and implementing activities that advance equity and social justice? • To what extent is there support or resistance for equity- and social justice–related activities or changes?	Evaluators in our sample investigated resistance to equity- and justice-focused programs, activities, or changes […] Examining resistance allows for a fuller understanding of the program and context, more nuanced insight into barriers to implementation, and deeper analysis of potential disconnects between equity- and social justice–focused program activities and desired outcomes. (Teasdale et al., 2024, pp. 18, 26)

Outcomes/ Impact	• What are the outcomes for minoritized populations? • Are there disparities in outcomes between members of minoritized and dominant groups? • To what extent have initial disparities decreased? • What changes are evident in systems, structures, policies, and other drivers of inequity?	Evaluators investigated the extent to which individuals or communities of color or Indigenous individuals or communities experienced or made progress toward positive outcomes. They also checked for disparities and whether pre-program disparities were reduced. (Teasdale et al., 2024, p. 14) How do outcomes differ across cultural (e.g., racial/ethnic, gender, language) groups? (short, intermediate, or longer term) To what extent are the burdens and benefits of the project distributed across different racial/ethnic and socioeconomic status groups? (short, intermediate, and longer term) (Thomas & Campbell, 2021, p. 319) How has the project produced changes to investment, policy, or practice that will enable change in infrastructure or scale? (Thomas & Campbell, 2021, p. 319) …[data] about the outcomes of the change initiative should at a minimum be disaggregated, so that differential effects by race, ethnicity, gender, language, or a myriad of other dimensions can be spotted and accounted for. (Dean-Coffey, 2018, p. 535) (An equity-focused evaluation) provides assessments of what works and what does not work to reduce inequity, and it highlights intended and unintended results for worst-off groups as well as the gaps between best-off, average and worst-off groups. (Segone, 2012, p. 7) Did the grantee's work affect race or intergroup relations, institutional changes, equity, socioeconomic status, or disparities of outcomes in the priority community? (Elam & Walker, 2021, p. 34)
Reach/ Access	• To what extent do members of minoritized groups have access and opportunity to participate? • What barriers do they encounter? • To what extent does the evaluand engage and retain members of minoritized groups? • To what extent does the evaluand focus on and reach	What environmental (e.g., transportation access) or contextual factors (e.g., community socioeconomic status, language barriers, unequal distribution of power, historic treatment of minorities or immigrant populations in the target community) are barriers to the participants' access to the project's services? (Thomas & Campbell, 2021, p. 317) How well does the project access hard-to-reach populations? Who missed out? (Thomas & Campbell, 2021, p. 317)

		individuals or communities with the most need?	Who cannot participate, and why not? (Mertens & Hopson, 2006, p. 45)
		• Are there disparities in access, participation, or retention between members of minoritized and dominant groups?	Are public and private service delivery systems reaching the worst-off groups? What are the main constraints on supply and demand? (Bamberger & Segone, 2011, p. 36)
		• Who did not or cannot participate?	
		• Who is not being served?	To what extent has this project increased diversity of participants? […] How and in what ways is the project ensuring that various populations have access to resources? (Boyce & Smith, 2021)
			…equity may be used in the evaluation as a criterion of merit of the evaluand. This may apply to program process (e.g., equity of access or experience with program services)… (Kirkhart, 2016, p. 116)
			Evaluators in our sample investigated how programs recruited participants and who programs sought to engage, as well as levels of program participation, retention, and attrition […] Evaluators also examined barriers to participation. (Teasdale et al., 2024, p. 17)
Relevance		• In what ways is the evaluand relevant to the needs and circumstances of minoritized populations?	How does the theory of change compare and interact with the cultural values, lifestyles, and worldview of clients and the surrounding community? (Thomas & Campbell, 2021, p. 317)
		• In what ways is the evaluand appropriate for the cultures, languages, and interests of minoritized populations?	To what extent was the project implemented […] culturally appropriately…? (Thomas & Campbell, 2021, p. 317)
		• How does it align with the cultural values, lifestyles, and worldviews of minoritized communities?	Jackson (1940a) framed the "advanced organizers" in evaluating African American education by indicating that "the needs of the group, individually, and collectively, must serve as a criterion if a true democracy is to be achieved" (Hood, 2001, p. 38)
			… a CRE approach is likely to include a guiding question on the extent to which the project's philosophy compares and interacts with the cultural values of its clients and the surrounding community (Thomas & Parsons, 2017, p. 17)
			What is the value of the intervention in relation to the needs of the worst-off groups…? (Bamberger & Segone, 2011, p. 35)

Replicability	• To what extent can the evaluand be replicated or scaled up to address inequities elsewhere? • How does the evaluand balance responsiveness to local context with aims for replication and scaling?	Will the strategy be more widely replicated or adapted? Is it likely to be scaled-up ("go to scale")? (Bamberger & Segone, 2011, p. 37) …evaluators should attend to the ways that context and culture constitute programs, examine how replication may or may not be appropriate for a given program, and identify what adaptations may be needed if program models are transferred across contexts. (Garibay & Teasdale, 2019, p. 100)
Resource Allocation / Resource Use	• To what extent are the resources allocated to the evaluand sufficient in comparison to the size and depth of historical inequities the evaluand aims to address? • To what extent does the evaluand direct its resources toward communities with the most need? • To what extent does the evaluand direct its resources to reduce disparities in access or outcomes between members of minoritized and dominant groups? • To what extent do the resources actually reach the communities in need? • To what extent are the costs appropriate compared with the equity gains?	To what extent is there appropriate and equitable use of resources? (Thomas & Campbell, 2021, p. 317) Does the programme use resources in the most economical manner to achieve expected equity-focused results? Are any other economic alternatives feasible? How cost-effective are the public systems for reaching worst-off groups? How do the costs of reaching the worst-off groups compare average costs and alternative ways of reaching them? (Bamberger & Segone, 2011, p. 36) How does resource distribution affect stakeholders' ability to benefit from the innovations? (Mertens & Hopson, 2006, p. 45) Accountability can be at the point of inputs. A community accountability project in Uganda reported in newspapers the amount of money that local schools were supposed to be receiving, making evident how much leakage was occurring (Reinikka & Svenson, 2005). (Rogers, 2016, p. 2016) Obviously a grantor who is awarding small grants of $20,000 per grantee cannot expect an agency to reduce teen pregnancy in a community by 50 percent. If the agency dedicated its entire budget to this cause, it would not have any effect on the systematic problems that cause the condition to exist. (Madison, 2000, p. 25)
Sustainability	• How likely is it the evaluand will have long-term benefits for minoritized populations through continuation of the evaluand or its outcomes? • How likely are those benefits to continue after the evaluand ends? • To what extent are additional benefits anticipated in the future?	Is the intervention and its impact on the worst-off groups likely to continue when external support is withdrawn? Are inequities between best-off and worst-off groups likely to increase, remain stable, or decrease when external support is withdrawn? (Bamberger & Segone, 2011, p. 37) Did the provision of services have a different long-term impact on various cultural groups after the conclusion of service delivery? (Elam & Walker, 2021, p. 31)

	• Do long-term benefits vary between members of minoritized and dominant groups?	To what extent are policy and infrastructure changes being made in a way that is likely to lead to sustained environmentally-focused place-based education that fits the cultural values of the community and is socially just? (Thomas & Parsons, 2017, p. 24)
Unintended Effects	• To what extent and in what ways does the evaluand cause harm or other unintended negative consequences to minoritized populations? • To what extent does the evaluand provide additional advantage to dominant communities or to those who are already advantaged?	[Evaluation] can monitor the gaps between the best-off and worst-off groups. It can watch for unintended consequences of well-intended programs that actually widen the gaps. (Kirkhart, 2016, 116) Equity-focused evaluations need to both anticipate unintended outcomes and impacts, as part of good risk management, and provide opportunities to notice and document outcomes and impacts that were unanticipated as well as unintended. (Rogers, 2016, p. 204) From this position, whether using data as evidence for injustice or illuminating the unintended outcomes that may cause harm to families impacted by a program so that an organization can recalibrate, advocacy is uncompromisingly speaking truth to power. (McBride et al, 2020, p. 124) [An equity-focused evaluation…] highlights intended and unintended results for worst-off groups. (Bamberger & Segone, 2011, p. 9) Were there unintended changes or consequences because of cultural or ethnic issues/context? (Elam & Walker, 2021, p. 25)

To generate the framework, we began by modifying the list of domains in Table 1. First, we removed the criterion of Equity / Social Justice from the list and shifted it to become the heading of a new column: "Equity / Social Justice Domain Applied as a Cross-Cutting Lens." We created two subcolumns, as explained below. Second, we added a new criteria domain, Openness, to the list, described below, based on our prior research (Avent et al., 2023, Teasdale et al., 2024).

Next, we located guidance and examples about how evaluators can advance equity and social justice through the lines of inquiry they pursue. We understood there was limited discussion of equity-focused criteria and lines of inquiry within the evaluation literature (Gates, Williamson, et al., 2022; Teasdale, 2022a; Teasdale, Strasser, et al.,

2023). Therefore, rather than reviewing the literature broadly, we focused specifically on articles, book chapters, and evaluation guides that (a) addressed social justice–oriented evaluation and (b) discussed evaluative lines of inquiry. We did not seek to conduct an exhaustive review or synthesis. Instead, our goal was to locate advice and illustrations to refine the framework. When we located relevant guidance and examples, we mapped each excerpt to the criteria domain it addressed. We have included an abbreviated set of guidance and examples in the Table 2 subcolumn labeled "Guidance and Examples from the Evaluation Literature." Finally, we drew on the guidance and examples to draft essential questions that encapsulate the application of the Equity / Social Justice criterion as a lens. We present these

questions in the "Essential Questions" subcolumn, mirroring and expanding the essential questions included in Table 1.

In the following sections, we summarize each of the 11 domains in Table 2 and describe how applying the Equity / Social Justice criterion as a lens can focus evaluative lines of inquiry. Then, we apply them to a hypothetical evaluation example to illustrate. Using a hypothetical enables us to illustrate each of the 11 domains and lines of inquiry. For a real-world evaluation, only some of those domains and lines of inquiry would be relevant, as discussed below.

Example Evaluation

Our example evaluation focuses on a hypothetical higher education program: Research Experience in Chemistry (REC). REC is based on real-world programs that engage and support undergraduates in conducting hands-on research in university science laboratories (National Academies of Sciences, Engineering, and Medicine, 2017). Our hypothetical REC takes place at a predominantly white, research-intensive university and provides full-time, paid summer research positions for students. REC aims to foster students' interest in chemistry research, develop their research skills, and increase the number of students who pursue graduate education in chemistry. REC includes a focus on engaging and supporting racially minoritized students.

Alignment

The Alignment domain addresses the ways in which an evaluand is consistent or coordinated with related programs or initiatives. Applying the Equity / Social Justice criterion as a lens can focus evaluative lines of inquiry on the extent to which an evaluand contributes to broader priorities or a longer-term strategy to advance equity and social justice. Lines of inquiry can also examine how the evaluand builds on or advances prior or current equity efforts.

To address alignment, our hypothetical evaluation might investigate coordination of REC activities with other efforts to engage and support racially minoritized undergraduates. For example, the evaluation might examine the extent to which REC goals are aligned with and advance campus-level goals for equity, diversity, and inclusion or broader efforts. In addition, some chemistry departments invest in training for faculty and staff that addresses equity and inclusion in pedagogy, mentoring, and lab culture and climate. The

evaluation might examine the selection process for REC labs, considering whether the program prioritizes those who have completed this type of training. The evaluation might also consider the ways in which REC connects racially minoritized students with other campus or departmental supports. These lines of inquiry could reveal how REC contributes to broader priorities for advancing racial equity and justice and coordinates with existing efforts—or remains siloed, duplicates existing activities, or works at cross-purposes.

Design/Implementation

The Design/Implementation domain focuses on the quality of the evaluand's conceptualization and execution, including the extent to which the evaluand fits the relevant context(s) and is implemented as planned. When applying the Equity / Social Justice criterion as a lens, evaluative inquiry can address the ways in which culture, context, and equity-focused processes or goals are addressed in the evaluand's development. In addition, lines of inquiry can examine whether and how the evaluand employs inclusive, equitable, culturally sustaining, or culturally relevant practices and whether and how activities reflect the values and perspectives of minoritized populations. More fundamentally, lines of inquiry can investigate the extent to which the evaluand addresses root causes of inequities and its underlying program theory reflects asset-based or deficit perspectives of minoritized populations.

For our hypothetical REC, the evaluation might consider whether and how REC scientists are trained in inclusive and equitable mentoring practices and the extent to which structures are in place to identify and address disparate treatment, bias, or discrimination. The evaluation might also examine the extent to which REC fosters community among racially minoritized participants to counter isolation and provide support and resources for navigating the predominantly white university context. More fundamentally, the evaluation might consider the extent to which REC takes the root causes of racial inequities in chemistry into account. For example, the evaluation might probe for deficit assumptions about racially minoritized students among faculty or discriminatory practices within the chemistry department or broader discipline. This could help program leaders address the structures, mindsets, and practices that drive inequities, in addition to addressing students' needs for experience with chemistry research.

Experience

The Experience domain addresses participants' experiences with the evaluand, such as the extent to which they are treated with respect, find activities to be rewarding or enjoyable, or feel satisfied with the evaluand. Applying the Equity / Social Justice criterion as a lens can focus evaluative lines of inquiry on the experiences of participants who are members of minoritized groups. Lines of inquiry can also investigate disparities in experiences between minoritized and dominant groups and identify contexts and actions that produce disparities.

Our hypothetical evaluation of REC might examine the experiences of racially minoritized students within their research labs and investigate any disparities in experiences among students of different racial identities. For example, the evaluation might address racially minoritized students' level of satisfaction with their research experiences, their sense of belonging and feeling valued within their labs, or the extent to which students feel they can bring their whole selves to REC. Further, the evaluation might investigate experiences of discrimination or bias and address lab cultures and climates. These lines of inquiry could reveal whether and how REC fosters respectful and rewarding experiences for racially minoritized students.

Openness

The Openness domain emerged through our recent research (Avent et al., 2023, Teasdale et al., 2024). Drawing on the organizational change literature, we define openness as willingness to support change, and positive attitude about potential consequences of change (Augustsson et al., 2017; Miller et al., 1994; Wanberg & Banas, 2000). This domain addresses openness among staff and leaders to design and implement evaluand activities and needed changes. Applying the Equity / Social Justice criterion as a lens can focus evaluative lines of inquiry on the extent to which evaluand staff and leaders are open to designing and implementing activities that advance equity and social justice. Lines of inquiry can investigate support for and resistance to equity- and social justice–related activities and changes.

Our example evaluation might address the extent to which faculty and staff demonstrate support for or resistance to REC's focus on engaging and supporting racially minoritized students. This might include openness to participating in equity-related training; providing inclusive and equitable mentoring; ensuring equitable opportunities, experiences, and outcomes during the program; monitoring equity of those opportunities, experiences, and outcomes; and making changes to foster greater equity. This might also include openness to addressing the root causes of racial inequities within the chemistry department and the broader discipline. These lines of inquiry could help to illuminate the program context, barriers to implementation, and factors that might limit achievement of desired outcomes.

Outcomes/Impact

The Outcomes/Impact domain addresses the extent to which the evaluand achieves its intended short-term or long-term outcomes, impact, or results. Applying the Equity / Social Justice criterion as a lens can focus evaluative lines of inquiry on the outcomes realized for minoritized populations, as well as outcome disparities between minoritized and dominant groups. In addition, lines of inquiry can investigate the extent to which initial disparities are reduced, maintained, or increased. Lines of inquiry can also address changes in structures, policies, and other drivers of inequity, in addition to or in place of individual-level outcomes.

Our example evaluation might investigate the extent to which the intended outcomes of REC are realized for racially minoritized students: increased interest in chemistry, research skills, and graduate school enrollment. The evaluation might also look for disparities in outcomes among students of different racial identities. In addition, the evaluation might examine disparities in interest or skills at the start of REC and whether disparities are reduced, maintained, or increased over the course of the program.

Reach/Access

The Reach/Access domain focuses on the extent to which the evaluand engages and retains intended participants. When applying the Equity / Social Justice criterion as a lens, inquiry can address the extent to which the evaluand engages and retains members of minoritized groups or populations with greatest need. Lines of inquiry can also examine the extent to which members of minoritized groups have access and opportunity to participate and identify barriers within and beyond the program. Further, lines of inquiry can focus on disparities in access, participation, or retention between minoritized and dominant groups and the drivers of

those disparities. Lines of inquiry can also address those in need who do not or cannot participate.

Our hypothetical evaluation might investigate the number of racially minoritized students who participate in REC, examine their attrition over time, and look for any disparities among students of different racial identities. The evaluation could then identify barriers to participation for racially minoritized students and factors driving differential participation or attrition. The evaluation might also consider racially minoritized students who do not participate in REC, especially those with high need or interest, and identify barriers that limit access and opportunity.

Relevance

The Relevance domain addresses the ways in which the evaluand's aims and activities are consistent with the needs, requirements, cultures, interests, or circumstances of participants or intended beneficiaries. Applying the Equity / Social Justice criterion as a lens can focus evaluative lines of inquiry on whether the evaluand fits the needs, priorities, values, interests, and circumstances of minoritized populations and is culturally and linguistically appropriate.

Our example evaluation might examine the fit between REC activities and opportunities and the needs, priorities, and circumstances of racially minoritized students. For example, the evaluation might consider the types of research conducted by participating labs and how well they match students' interests, values, and goals. The evaluation might also consider how well the program model of full-time summer employment fits with students' family and employment responsibilities. More broadly, the evaluation might identify barriers racially minoritized students experience in pursuing graduate education in chemistry, resources students leverage to address them, and barriers that remain unaddressed. This could reveal whether racially minoritized students need the research experience REC provides or other types of support.

Replicability

The Replicability domain focuses on the extent to which the evaluand—or its components, activities, or underlying model—can be duplicated or adapted to another context. When applying the Equity / Social Justice criterion as a lens, evaluative inquiry can address how the evaluand can be replicated or scaled up to address inequities elsewhere and how

to balance responsiveness to local context with aims for replication and scaling.

Our hypothetical evaluation might identify the REC components or activities that could be duplicated or adapted to support racially minoritized chemistry undergraduates at other universities. For example, the evaluation could determine which elements of mentor training or community-building activities best support racially minoritized students in REC, as well as how those elements have been tailored to REC's unique student body and departmental and university contexts. By documenting key program elements and their associated context, the evaluation could aid other universities in replicating REC while being responsive to their local contexts.

Resource Allocation / Resource Use

The Resource Allocation / Resource Use domain addresses the extent to which funding, personnel, and materials are sufficient to implement the evaluand and are used economically.

Applying the Equity / Social Justice criterion as a lens with this domain can focus evaluative lines of inquiry on the extent to which allocated resources are sufficient in comparison to the size and depth of historical inequities and injustices the evaluand aims to address. In addition, lines of inquiry can address the extent to which resources are directed to communities with the most need or to reduce disparities between minoritized and dominant groups. Evaluation can also investigate the extent to which allocated resources actually reach priority communities and whether costs are appropriate when compared with equity gains.

Our example evaluation might assess the size of the investment in REC and the proportion of expenditures that directly support racially minoritized participants. This could reveal the extent to which resources are commensurate with the size and depth of historical inequities within the chemistry department, university, and broader discipline, and the likelihood that the invested resources will contribute to meaningful change.

Sustainability

The Sustainability domain focuses on the extent to which the evaluand has—or is likely to have—long-term benefits. This can be accomplished through outcomes that persist over time, continuation of the evaluand itself, or environmentally sustainable activities or processes. When applying the Equity / Social Justice criterion as a lens to this domain, evaluative inquiry can address the extent to which

the evaluand cultivates long-term benefit for minoritized populations or whether additional positive results are anticipated in the future. Lines of inquiry can also examine variation in long-term benefits between minoritized and dominant groups.

The hypothetical evaluation of REC might examine pathways and experiences of racially minoritized participants after they complete REC. For example, after documenting the program's intended short- and mid-term outcomes—students' interest in chemistry research, research skills, and graduate school enrollment—the evaluation could examine skills participants gain in graduate school or completion rates of chemistry graduate degrees to understand the longer-term outcomes of REC. The evaluation could also examine students' career pathways, such as their experiences pursuing positions in academia and industry. The evaluation could compare the pathways and experiences of students of different racial identities to illuminate variations and disparities.

Unintended Effects

The Unintended Effects domain addresses side effects of the evaluand, including both negative and positive unintended consequences. Applying the Equity / Social Justice criterion as a lens with this domain can focus evaluative lines of inquiry on whether the evaluand causes harm or other unintended negative consequences to minoritized populations, as well as whether it provides additional advantage to dominant communities.

To do this, our hypothetical evaluation might look for instances in which REC activities, processes, or structures cause harm to racially minoritized program participants. This would require open-ended inquiry that centers students' experiences and voices and is sensitive to experiences and consequences of differential treatment, bias, and discrimination. The evaluation might also look for harm and negative consequences for racially minoritized students who are not selected for the program or are unable to participate—such as missed opportunities to conduct research or develop professional networks—or consider how investments in REC might inadvertently decrease funding for other programs. Finally, the evaluation might look for additional advantage that accrues to white REC participants. For example, the evaluation might consider whether REC enhances research skills and professional networks for students who have already had multiple opportunities and support to develop them. Thus, REC could inadvertently

provide white students with further advantage over racially minoritized students who have not been afforded similar opportunities and support.

Using the Framework

The framework outlines possibilities available to evaluators for investigating evaluands' contribution to advancing equity and social justice or, conversely, to reinforcing inequity and injustice. Evaluators and constituents can use the framework as a thinking tool to map the broad categories of quality that may be relevant for a specific evaluand and context. These lines of inquiry can be combined with those described in the larger criteria model, which are not explicitly focused on equity and social justice yet can contribute in important ways to understanding the evaluand and its context(s) (Teasdale et al., 2024). A single evaluation cannot investigate all possible definitions of quality, and not all criteria are relevant for every evaluand. Thus, evaluators and constituents can use the framework to choose the most salient domains. Then they can plan lines of inquiry to address them and operationalize each broad category of quality to identify appropriate equity- and justice-focused indicators and metrics.

For example, evaluators can use the essential questions in Table 2 to surface constituents' varying values about the characteristics or results that matter most and make those values explicit within and across groups (Gates et al., 2024; Greene, 2012; Gullickson, 2020; Gullickson & Hannum, 2019; Teasdale, Pitts, et al., 2023). The framework provides concepts and language for engaging constituents in a reflective process to examine varying values about evaluand quality (Schwandt & Gates, 2021; Tovey & Archibald, 2023) and negotiate differing perspectives on equity and social justice (Avent et al., 2023). This can take place throughout the evaluation to illuminate criteria that emerge or change over time (Teasdale, Pitts, et al., 2023). In addition, the framework can be used to challenge or expand preset lines of inquiry that do not adequately address equity and social justice (Gates, Williamson, et al., 2022; Teasdale et al., 2024).

The framework can also be used to explicate the relationship between values and evaluation design. A particular evaluation could be organized around multiple possible lines of inquiry, guided by different domains that reflect different values. Evaluators often combine multiple lines of inquiry into a single study to address complex evaluands or a range of values (House & Howe, 1999; Teasdale, 2022a; Teasdale, Strasser, et al., 2023).

Articulating the criterion that directs each line of inquiry can clarify the values that underpin each evaluation question, method, and conclusion. The domains can also be used to organize reporting and provide clear, complete documentation of the evaluation focus, process, and findings and the values that guided it (Teasdale, Moore, et al., in press). Evaluation educators, coaches, and mentors can use the framework to guide novice evaluators in learning to conceptualize, name, discuss, and investigate lines of inquiry that address equity and social justice. Scholars can use the framework to surface the values that underpin evaluations and prescribe criteria for specific types of evaluands, contexts, or evaluation approaches.

This framework, like the criteria model on which it is built, is intended to be dynamic. The framework can be used as the basis for future research and then refined to incorporate new understandings. Specifically, we see the need for inquiry to better understand how the criteria in the framework are used in practice. Future research is also needed to identify the sources of those criteria, the processes used to select them, and how evaluators navigate differing values and conceptualizations related to equity and social justice (Teasdale et al., 2024). This requires research with a range of constituents—program participants, community members, program leaders, and funders—to understand their values and how they define equity and social justice for their evaluands and contexts (Avent et al., 2023; Boyce et al., 2023). Further research is also needed to understand how evaluators' identities and lived experiences shape the criteria and lines of inquiry they pursue (Teasdale et al., 2024). This scholarship could contribute to continued development of the framework and further support evaluators in investigating equity and social justice.

Conclusion

Evaluators have a professional and ethical responsibility to contribute to the "advancement of an equitable and just society" (American Evaluation Association, 2018b). In this article, we have presented a framework for investigating equity and social justice within programs through the criteria and lines of inquiry evaluators pursue. The framework outlines 11 domains in which a program's contribution to equity and social justice might be examined. We have discussed an example evaluation to illustrate each domain and concluded by describing how the framework can be used to advance equity and social justice in evaluation practice, education, and scholarship.

References

American Evaluation Association. (2018a). *AEA evaluator competencies*. https://www.eval.org/About/Competencies-Standards/AEA-Evaluator-Competencies

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

American Evaluation Association. (n.d.). *Dialogues on race and class*. https://www.eval.org/Events/Dialogues-on-Race-and-Class

American Medical Association. (2021). Advancing health equity: A guide to language, narrative and concepts. https://www.ama-assn.org/about/ama-center-health-equity/advancing-health-equity-guide-language-narrative-and-concepts-0

Armytage, L. (2011). Evaluating aid: An adolescent domain of practice. *Evaluation*, *17*(3), 261–276. https://doi.org/10.1177/1356389011410518

Augustsson, H., Richter, A., Hasson, H., & von Thiele Schwarz, U. (2017). The need for dual openness to change: A longitudinal study evaluating the impact of employees' openness to organizational change content and process on intervention outcomes. *The Journal of Applied Behavioral Science*, *53*(3), 349–368. https://doi.org/10.1177/0021886317691930

Avent, C. M. (2025). The criticality of language: Exploring STEM education evaluators conceptualizations of equity, diversity, and inclusion and the influence on their roles and practice. *Evaluation and Program Planning*, *108*. https://doi.org/10.1016/j.evalprogplan.2024.102511

Avent, C., Teasdale, R. M., Yan, X., Serrano Abreu, M. B., & Moore, C. (2023). *Towards racial equity: Navigating diverse conceptualizations and perspectives in social justice-oriented evaluation practice* [Manuscript submitted for publication].

Bamberger, M., & Segone, M. (2011). *How to design and manage equity-focused evaluations*. https://agora.unicef.org/course/info.php?id=1238

Benitez, M., Jr. (2010). Resituating culture centers within a social justice framework: Is there room for examining Whiteness? In L. D. Patton (Ed.), *Culture centers in higher education:*

Perspectives on identity, theory, and practice (pp. 119–134). Stylus.

Bledsoe, K. (2021). Evaluation's identity: Who are we, why are we here, and what does our future look like? *Evaluation Matters—He Take Tō Te Aromatawai*, *7*, 6–29. https://doi.org/10.18296/em.0064

Bowen, M. L., & Tillman, A. S. (2015). Developing culturally responsive surveys: Lessons in development, implementation, and analysis from Brazil's African descent communities. *American Journal of Evaluation*, *36*(1), 25–41. https://doi.org/10.1177/1098214014539831

Boyce, A., & Smith, T. (2021). *Diversity, equity, and inclusion in evaluation: Definitions, evaluation questions, indicators, and data collection methods.* https://evalu-ate.org/guide/edi-in-evaluation/

Boyce, A. S., Tovey, T. L. S., Onwuka, O., Moller, J. R., Clark, T., & Smith, A. (2023). Exploring NSF-funded evaluators' and principal investigators' definitions and measurement of diversity, equity, and inclusion. *American Journal of Evaluation*, *44*(1), 50–73. https://doi.org/10.1177/10982140221108662

Canadian Evaluation Society. (n.d.). *2019–2024 Canadian Evaluation Society strategic plan.* https://evaluationcanada.ca/client_assets/PDFs/CES%20Strategic%20Plan%20EN.pdf

Centre for Excellence for Evaluation. (2015). *Evaluation of the 2009 Policy on Evaluation.* https://www.canada.ca/en/treasury-board-secretariat/services/audit-evaluation/centre-excellence-evaluation/evaluation-2009-policy-evaluation.html

Chianca, T. (2008). The OECD/DAC criteria for international development evaluations: An assessment and ideas for improvement. *Journal of MultiDisciplinary Evaluation, 5*(9), 41–51. https://doi.org/10.56645/jmde.v5i9.167

Chilisa, B., & Bowman, N. R. (2023). Special issue editors' introductory note: The why and how of the decolonization discourse. *Journal of MultiDisciplinary Evaluation*, *19*(44), 2–10. https://doi.org/https://doi.org/10.56645/jmde.v19i44.919

Chilisa, B., Major, T. E., Goathobogwe, M., & Mokgolodi, H. (2016). Decolonizing and indigenizing evaluation practice in Africa: Toward African relational evaluation approaches. *Canadian Journal of Program Evaluation*, *30*(3), 313–328. https://doi.org/10.3138/cjpe.30.3.05

Cram, F. (2016). Lessons on decolonizing evaluation from Kaupapa Māori evaluation. *Canadian Journal of Program Evaluation*, *30*(3), 296–312. https://doi.org/10.3138/cjpe.30.3.04

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

Davidson, E. J. (2005a). Criteria. In S. Mathison (Ed.), *Encyclopedia of Evaluation* (pp. 91–92). Sage.

Davidson, E. J. (2005b). *Evaluation methodology basics: The nuts and bolts of sound evaluation.* Sage.

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. https://doi.org/10.1177/1098214018778533

Dickinson, P., & Adams, J. (2017). Values in evaluation: The use of rubrics. *Evaluation and Program Planning*, *65*, 113–116. https://doi.org/10.1016/j.evalprogplan.2017.07.005

Dumaine, F. (2012). When one must go: The Canadian experience with strategic review and judging program value. *New Directions for Evaluation*, *133*, 65–75. https://doi.org/10.1002/ev.20007

Elam, P., & Walker, W. (2021). *Considerations for conducting evaluation using a culturally responsive and racial equity lens.* https://mphi.org/wp-content/uploads/2022/05/Considerations-for-Conducting-Evaluation-Using-a-Culturally-Responsive-and-Racial-Equity-Lens.pdf

Fournier, D. M. (1995). Establishing evaluative conclusions: A distinction between general and working logic. *New Directions for Evaluation*, 68, 15–32. https://doi.org/10.1002/ev.1017

Garibay, C., & Teasdale, R. M. (2019). Equity and evaluation in informal STEM education. *New Directions for Evaluation*, *161*, 87–106. https://doi.org/10.1002/ev.20352

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

Gates, E. F., Teasdale, R. M., Shim, C., & Hubacz, H. (2024). Whose and what values? Advancing and illustrating explicit specification of evaluative criteria in education. *Studies in Educational Evaluation*, *81*, 101335. https://doi.org/10.1016/j.stueduc.2024.101335

Gates, E. F., Williamson, E., Madres, J., Hall, J. N., & Alvarez, K. B. (2022). Equity "on the

sideline": A mixed methods study of New England evaluation practice in 2020. *Journal of MultiDisciplinary Evaluation*, *18*(42). https://doi.org/10.56645/jmde.v18i42.715

Giacomini, M., & Hurley, J. (2008). Issues in evaluating equity. In L. Potvin, D. V. McQueen, M. Hall, L. de Salazar, L. M. Anderson, & Z. M. Hartz (Eds.), *Health promotion evaluation practices in the Americas* (pp. 285–298). Springer.

Greene, J. C. (2006). Evaluation, democracy, and social change. In I. Shaw, J. C. Greene, & M. M. Mark (Eds.), *The SAGE Handbook of Evaluation* (pp. 118–140). Sage.

Greene, J. C. (2012). Values-engaged evaluations. In M. Segone (Ed.), *Evaluation for equitable development results* (pp. 192–206). UNICEF. http://www.clear-la.cide.edu/sites/default/files/Evaluation_for_equitable%20results_web.pdf

Greene, J. C., Boyce, A. S., & Ahn, J. (2011). *A values-engaged, educative approach for evaluating education programs: A guidebook for practice*. University of Illinois at Urbana-Champaign. http://comm.eval.org/communities/community-home/librarydocuments/viewdocument?DocumentKey=f3c734c0-8166-4ba4-9808-a07e05294583

Gullickson, A. M. (2020). The whole elephant: Defining evaluation. *Evaluation and Program Planning*, *79*, 101787. https://doi.org/10.1016/j.evalprogplan.2020.101787

Gullickson, A. M., & Hannum, K. M. (2019). Making values explicit in evaluation practice. *Evaluation Journal of Australasia*, *19*(4), 162–178. https://doi.10.1177/1035719X19893892

Hall, J. N. (2020a). *Focus groups: Culturally responsive approaches for qualitative inquiry and program evaluation*. Myers Education.

Hall, J. N. (2020b). 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. (2018). Evaluation's race problem in the U.S.: 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, M. E. (2020). Blest be the ties that bind. *New Directions for Evaluation*, *166*, 13–22. https://doi.org/10.1002/ev.20414

Hood, S. (2001). Nobody knows my name: In praise of African American evaluators who were responsive. *New Directions for Evaluation*, *92*, 31-44. https://doi.org/10.1002/ev.33

Hood, D. W., & Cassaro, D. A. (2002). Feminist evaluation and the inclusion of difference. *New Directions for Evaluation*, *96*, 27–40. https://doi.org/10.1002/ev.64

Hood, S., Hopson, R. K., & Kirkhart, K. E. (2015). Culturally responsive evaluation: Theory, practice, and future implications. In K. E. Newcomer, H. P. Hatry, & J. S. Wholey (Eds.), *Handbook of practical program evaluation* (4th ed., pp. 281–317). Jossey-Bass.

Hopson, R., & Cram, F. (2018). Tackling wicked problems in complex evaluation ecologies. In R. Hopson & F. Cram (Eds.), *Tackling wicked problems in complex ecologies: The role of evaluation* (pp. 3–24). Stanford University Press.

House, E. R. (2015). *Evaluating: Values, biases, and practical wisdom*. Information Age.

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

Hurteau, M., Houle, S., & Mongiat, S. (2009). How legitimate and justified are judgments in program evaluation?. *Evaluation*, *15*(3), 307–319. https://doi.org/10.1177/1356389009105883

Jacobson, M. R., Azzam, T., & Baez, J. (2012). The nature and frequency of inclusion of people with disabilities in program evaluation. *American Journal of Evaluation*, *34*(1), 23–44. https://doi.org/10.1177/1098214012461558

Jordan, L. S., & Hall, J. N. (2023). Framing anticolonialism in evaluation: Bridging decolonizing methodologies and culturally responsive evaluation. *Journal of MultiDisciplinary Evaluation*, *19*(44), 102–116. https://doi.org/10.56645/jmde.v19i44.769

King, J., McKegg, K., Oakden, J., & Wehipeihana, N. (2013). Evaluative rubrics: A method for surfacing values and improving the credibility of evaluation. *Journal of MultiDisciplinary Evaluation*, *9*(21), 11–20. https://doi.org/10.56645/jmde.v9i21.374

Kirkhart, K. E. (2016). Equity, privilege, and validity: Traveling companions or strange bedfellows? In S. I. Donaldson & R. Picciotto (Eds.), *Evaluation for an equitable society* (pp. 109–131). Information Age.

Madison, A. (1992). Primary inclusion of culturally diverse minority program participants in the evaluation process. *New Directions for Evaluation*, *53*, 35–43. https://doi.org/10.1002/ev.1599

Madison, A. M. (2000). Language in defining social problems and in evaluating social programs. *New Directions for Evaluation*, *86*, 17–28. https://doi.org/10.1002/ev.1169

Manswell-Butty, J. L., Reid, M. D., LaPoint, V. (2004). A culturally responsive evaluation approach applied to the talent development school-to-career intervention program. *New Directions for Evaluation*, *101*, 37–47. https://doi.org/10.1002/ev.106

Mark, M. M., Henry, G. T., & Julnes, G. (2000). *Evaluation: An integrated framework for understanding, guiding, and improving policies and programs*. Jossey-Bass.

Martens, K. S. R. (2018). Rubrics in program evaluation. *Evaluation Journal of Australasia*, *18*(1), 21–44. https://doi.org/10.1177/1035719X17753961

McBride, D., Casillas, W., & LoPiccolo, J. (2020). Inciting social change through evaluation. *New Directions for Evaluation*, *166*, 119–127. https://doi.org/10.1002/ev.20405

Mertens, D. M. (1999). Inclusive evaluation: Implications of transformative theory for evaluation. *American Journal of Evaluation*, *20*(1), 1–14. https://doi.org/10.1177/109821409902000102

Mertens, D. M., & Hopson, R. K. (2006). Advancing evaluation of STEM efforts through attention to diversity and culture. *New Directions for Evaluation*, *109*, 35–51. https://doi.org/10.1002/ev.177

Mertens, D. M., & Wilson, A. T. (2019). *Program evaluation theory and practice: A comprehensive guide* (2nd ed.). Guilford.

Miller, R. L. (2018). Hiding in plain sight: On culturally responsive evaluation and LGBT communities of colour. *Evaluation Matters— He Take Tō Te Aromatawai*, *4*, 5–34. https://doi.org/10.18296/em.0027

Miller, V. D., Johnson, J. R., & Grau, J. (1994). Antecedents to willingness to participate in a planned organizational change. *Journal of Applied Communication Research*, *22*, 59–80. https://doi.org/10.1080/00909889409365387

Montrosse-Moorhead, B. (2022, October 13). *Evaluation synthesis: What is it? How do we teach it to others?* [Video]. https://vimeo.com/showcase/10952695/video/762437455

National Academies of Sciences, Engineering, and Medicine. (2017). *Undergraduate research experiences for STEM students: Successes, challenges, and opportunities*. National

Academies Press. https://doi.org/https://doi.org/10.17226/24622

Neubauer, L. C., McBride, D., Guajardo, A. D., Casillas, W. D., & Hall, M. E. (2020). Examining issues facing communities of color today: The role of evaluation to incite change. *New Directions for Evaluation*, *166*, 7–11. https://doi.org/10.1002/ev.20406

Organisation for Economic Co-operation and Development. (1991). *Principles for evaluation of development assistance*. http://www.oecd.org/dac/evaluation/50584880.pdf

Organisation for Economic Co-operation and Development. (1999). *Guidance for evaluating humanitarian assistance in complex emergencies*. https://www.oecd.org/dac/evaluation/2667294.pdf

Organisation for Economic Co-operation and Development. (2002). *Glossary of key terms in evaluation and results based management*. http://www.oecd.org/dac/evaluation/2754804.pdf

Organisation for Economic Co-operation and Development. (2018). *DAC criteria for evaluating development assistance*. http://www.oecd.org/dac/evaluation/daccriteriaforevaluatingdevelopmentassistance.htm

Organisation for Economic Co-operation and Development. (2019). *Better criteria for better evaluation: Revised evaluation criteria definitions and principles for use*. https://www.oecd.org/dac/evaluation/revised-evaluation-criteria-dec-2019.pdf

Ozeki, S., Coryn, C. L. S., & Schröter, D. C. (2019). Evaluation logic in practice: Findings from two empirical investigations of American Evaluation Association members. *Evaluation and Program Planning*, *76*, 101681. https://doi.org/10.1016/j.evalprogplan.2019.101681

Patton, M. Q. (2021). Evaluation criteria for evaluating transformation: Implications for the coronavirus pandemic and the global climate emergency. *American Journal of Evaluation*, *42*(1), 53-89. https://doi.org/10.1177/1098214020933689

Phillips, G. I., Felt, D., Perez-Bill, E., Ruprecht, M. M., Glenn, E. E., Linderman, P., & Miller, R. L. (2023). Transforming the paradigm for LGBTQ+ evaluation: Advancing a praxis of LGBTQ+ inclusion and liberation in evaluation. *American Journal of Evaluation*, *44*(1), 7–28. https://doi.org/10.1177/10982140211067206

Rogers, P. J. (2016). Understanding and supporting equity: Implications of methodological and procedural choices in equity-focused evaluations. In S. I. Donaldson & R. Picciotto (Eds.), *Evaluation for an equitable society* (pp. 199–216). Information Age.

Roorda, M., & Gullickson, A. M. (2019). Developing evaluation criteria using an ethical lens. *Evaluation Journal of Australasia*, *19*(4), 179–194. https://doi.org/10.1177/1035719X19891991

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

Sadler, D. R. (1985). The origins and functions of evaluative criteria. *Educational Theory*, *35*(3), 285–297. https://doi.org/10.1111/j.1741-5446.1985.00285.x

Schwandt, T. A. (2015). *Evaluation foundations revisited: Cultivating a life of the mind for practice*. Stanford University Press.

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

Scriven, M. (2000). Evaluation ideologies. In D. L. Stufflebeam, G. F. Madaus, & T. Kellaghan (Eds.), *Evaluation models: Viewpoints on educational and human services evaluation* (2nd ed., pp. 249–278). Kluwer.

Scriven, M. (2007). The logic of evaluation. In H. V. Hansen, C. W. Tindale, J. A. Blair, & R. H. Johnson (Eds.), *Proceedings of the 7th Ontario Society for the Study of Argumentation Conference.* https://scholar.uwindsor.ca/ossaarchive/OSSA7/papersandcommentaries/138

Scriven, M. (2012). The logic of valuing. *New Directions for Evaluation*, *133*, 17–28. https://doi.org/10.1002/ev.20003

Segone, M. (2012). Evaluation to accelerate progress towards equity, social justice and human rights. In M. Segone (Ed.), *Evaluation for equitable development results* (pp. 2–12). UNICEF. http://www.clear-la.cide.edu/sites/default/files/Evaluation_for_equitable%20results_web.pdf

Shipman, S. (2012). The role of context in valuing federal programs. *New Directions for Evaluation*, *133*, 53–63. https://doi.org/10.1002/ev.20006

Sielbeck-Bowen, K. A., Brisolara, S., Seigart, D., Tischler, C., & Whitmore, E. (2002). Exploring feminist evaluation: The ground from which we rise. *New Directions for Evaluation*, *96*, 3–8. https://doi.org/10.1002/ev.62

Stake, R., Migotsky, C., Davis, R., Cisneros, E. J., Depaul, G., Dunbar, C., Farmer, R., Feltovich, J., Johnson, E., Williams, B., Zuria, M., &

Chaves, I. (1997). The evolving syntheses of program value. *Evaluation Practice*, *18*(2), 89. https://doi.org/10.1016/S0886-1633(97)90015-5

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

Teasdale, R. M. (2022a). How do you define success? Evaluative criteria for informal STEM education. *Visitor Studies*, *25*(2), 163–184. https://doi.org/10.1080/10645578.2022.2056397

Teasdale, R. M. (2022b). Representing the values of program participants: Endogenous evaluative criteria. *Evaluation and Program Planning*, *94*, 102123. https://doi.org/10.1016/j.evalprogplan.2022.102123

Teasdale, R. M., Avent, C. M., Moore, C., Serrano Abreu, M. B., & Yan, X. (2024). "It has to be the common thread": Weaving attention to racial equity and justice across lines of inquiry and evaluative criteria. *American Journal of Evaluation.* Advanced online publication. https://doi.org/10.1177/10982140241260069

Teasdale, R. M., Moore, C., & Strasser, M. (in press). *Reporting lines of inquiry: Documenting evaluations and making values explicit. Canadian Journal of Program Evaluation.*

Teasdale, R. M., Pitts, R. T., Gates, E. F., & Shim, C. (2023). Teaching specification of evaluative criteria: A guide for evaluation education. *New Directions for Evaluation*, *117*, 31–37. https://doi.org/10.1002/ev.20546

Teasdale, R. M., Strasser, M., Moore, C., & Graham, K. (2023). Evaluative criteria in practice: Findings from an analysis of evaluations published in Evaluation and Program Planning. *Evaluation and Program Planning*, *97*, 102226. https://doi.org/10.1016/j.evalprogplan.2023.102226

Thomas, V. G. (2004). Building a contextually responsive evaluation framework: Lessons from working with urban school interventions. *New Directions for Evaluation 101*, 3–23. https://doi.org/10.1002/ev.104

Thomas, V. G., & Campbell, P. B. (2021). *Evaluation in today's world: Respecting diversity, improving quality, and promoting usability*. Sage.

Thomas, V. G., Madison, A., Rockcliffe, F., DeLaine, K., & McDonald Lowe, S. (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

Thomas, V. G., & Parsons, B. A. (2017). Culturally responsive evaluation meets systems-oriented evaluation. *American Journal of Evaluation, 38*(1), 7–28. https://doi.org/10.1177/1098214016644069

Tovey, T. L. S., & Archibald, T. (2023). The relationship between reflective practice, evaluative thinking, and practical wisdom. In M. Hurteau & T. Archibald (Eds.), *Practical wisdom for an ethical evaluation practice.* Information Age.

Treasury Board of Canada Secretariat. (2012). *Policy on evaluation.* https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=15024

Treasury Board of Canada Secretariat. (2016). *Policy on results.* https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=31300

United States General Accounting Office. (1988). *Children's programs: A comparative evaluation framework and five illustrations.* https://www.gao.gov/products/136936

Waapalaneexkweew. (2018). Looking backward but moving forward: Honoring the sacred and asserting the sovereign in Indigenous evaluation. *American Journal of Evaluation, 39*(4), 543–568. https://doi.org/10.1177/1098214018790412

Wanberg, C. R., & Banas, J. T. (2000). Predictors and outcomes of openness to changes in a reorganizing workplace. *Journal of Applied Psychology, 85*(1), 132–142. https://doi.org/10.1037/0021-9010.85.1.132

Yarbrough, D. B., Shulha, L. M., & Hopson, R. K. (2011). *The program evaluation standards: A guide for evaluators and evaluation users* (3rd ed.). Sage.