A culturally responsive evaluation framework and its application in environmental education

Kelley C. Anderson [a,*], Marc J. Stern [a], Robert B. Powell [b], Ashley A. Dayer [c], Thomas G. Archibald [d]

[a] *Department of Forest Resources and Environmental Conservation, Virginia Tech, Blacksburg, VA, USA*
[b] *Department of Parks, Recreation, and Tourism Management, Department of Forestry and Natural Resources, Clemson University, Clemson, SC, USA*
[c] *Department of Fish and Wildlife Conservation, Virginia Tech, Blacksburg, VA, USA*
[d] *Department of Agriculture, Leadership, and Community Education, Virginia Tech, Blacksburg, VA, USA*

ABSTRACT

Despite the many challenges inherent in conducting high-quality evaluations in the field of environmental education (EE), there is a growing recognition of the importance of evaluation, not only to gauge program success, but also to use evaluation results to improve programming, support organizational learning, and ensure programs are meeting the needs of diverse audiences. The challenges to conducting high-quality evaluations are exacerbated by historical issues of inequity and systemic racism that are pervasive in the United States and globally. We reviewed the literature on culturally responsive approaches to evaluation to propose a culturally responsive evaluation framework and consider its application in EE. This framework helps EE organizations and evaluators consider how issues of race, power, privilege, and inequity influence the evaluation process and the validity of evaluation findings. Implementing this framework may be resource-intensive, but it has the potential to improve evaluation processes and produce actionable results to further address issues of diversity, equity, and inclusion in the field of EE. We call for organizations that support EE efforts to recognize the importance of this approach, and provide adequate resources to encourage its implementation.

1. Introduction

Despite the many challenges inherent in conducting high-quality evaluations in the field of EE, there is a growing recognition of the importance of program evaluation, not only to gauge program success, but also to use evaluation results to improve programming, support organizational learning (Keene & Blumstein, 2010; Monroe, 2010) and ensure that programs are meeting the needs of diverse audiences (Ardoin, Clark, & Kelsey, 2013; Wojcik, Biedenweg, McConnell, & Iyer, 2013). These challenges include the heterogeneity of the field, which makes standard methods difficult to apply across different programs; a lack of clear programmatic goals and objectives, often stemming from an incomplete understanding of program logic and the limited time available to design EE programs and their evaluations; a lack of appropriate funding and capacity needed to produce rigorous and valid evaluations; and a common perception of evaluation as an accountability tool that could threaten program reputation and future funding opportunities

(Carleton-Hug & Hug, 2010). Additionally, the diversifying demographics of the United States and the growing belief that EE should be equitably distributed and inclusive in its content and approach have highlighted a critical need for EE organizations to examine the assumptions embedded within their programs and reconsider the needs and wants of more diverse audiences when designing and implementing programs. Further, the historical foundations of the environmental field stem from Eurocentric worldviews, which can perpetuate disparities in attendance, participation, and the effectiveness of EE programs between White people and BIPOC (Black, Indigenous, and People of Color) (McLean, 2013; Finney, 2014; Warren, Roberts, Breunig, & Alvarez, 2014). Barriers to participation and engagement in EE, such as content that does not align with participants' worldviews and experiences, fear of discrimination in the outdoors, and accessibility constraints, can unintentionally yet disproportionately exclude people of color (Pease, 2015; Roberts, 2007; Warren et al., 2014). A culturally responsive approach to evaluating EE programs can help to address both the need

---

for systematic program evaluation and the need to critically examine the underlying assumptions and values embedded in current programs to promote diversity, equity, and inclusion (DEI) within the field.

To better understand the current state of evaluation in EE, we present a brief review of the EE evaluation literature and highlight areas for future growth. We then describe a culturally responsive approach to evaluation that addresses issues of race, power, privilege, and inequity and provide insights into how this framework can be applied to the field of EE. We conclude with a call for organizations that implement and/or fund EE programs and evaluations to recognize the importance of this approach and the need to allocate adequate resources for its implementation. We also call on researchers to conduct case studies on culturally responsive evaluation processes in action and its application in EE to test and enhance this proposed framework.

1.1. The State of Evaluation in EE

The Belgrade Charter (1975) and the Tbilisi Declaration (1977) paved a clear path for the field of EE by defining overarching goals for the discipline, including influencing knowledge, attitudes, and behaviors pertaining to the environment and building critical thinking skills to mitigate current and future environmental issues (UNESCO, 1977). To determine if these goals are being met, program evaluations in EE have most commonly been summative accounts of participant satisfaction and outcome achievement in the areas of knowledge, attitudes, behaviors, and skills (Ardoin, Bowers, Roth, & Holthuis, 2018; Bourke, Buskist, & Herron, 2014; Stern, Powell, & Hill, 2014). Regardless of which specific outcomes are of importance to EE organizations, the current literature reveals a heavy focus on outcome assessments that do not necessarily reveal how, why, or for whom programmatic outcomes are achieved (Stern et al., 2014). The ubiquity of outcome evaluations in EE may stem from the origins of evaluation in the field, which rose from formal education (Stake, 1967). These evaluations often prioritized the needs and expectations of funding agencies and were often operationalized as an accountability tool (Carleton-Hug & Hug, 2010). Many have critiqued this approach, because it tends to hamper the practical use of evaluation results for improving EE (Carman & Fredericks, 2008; Hoole & Patterson, 2008; Patton, 2003).

Realizing the missed opportunities that an accountability-focused evaluation approach created, Patton (2003) proposed a utilization-focused evaluation approach that emphasizes designing evaluations for utility and use of the results by program providers. This approach became popular in the field of EE because it gave more power to program directors and staff to determine which evaluation processes and products would be of most use to them (Crohn & Birnbaum, 2010; Greene, 2010; Powell, Stern, & Ardoin, 2006). This shift in thinking allowed for evaluation processes and methodologies to be tailored to specific programs and to provide evaluation findings that were salient to intended users, leading to an increased use of the findings to make empirically-based decisions about current programming.

Patton (2002) also proposed incorporating qualitative and mixed-methods in evaluations, which has only slowly begun to appear in the EE literature (Stern et al., 2014). Historically, the most common methods used to evaluate outcomes in EE have included quantitative designs using surveys administered to program participants, teachers, and chaperones (Leeming, Dwyer, Porter, & Cobern, 1993; Rickinson, 2001; Stern et al., 2014). This *etic* approach to evaluation, or evaluating to understand changes in programmatic outcomes from a neutral observer's perspective, can miss opportunities to understand program impacts from the viewpoints and voices of the participants (Patton, 2002). In a recent literature review including 119 program evaluations conducted on EE programs with school-aged participants, Ardoin and colleagues (2018) found that 82% implemented a standardized quantitative measure, 29% used interviews to collect data, 19% used observations, and 13% reviewed students work such as journals and writing prompts. While the review was limited to the peer-reviewed literature, it

suggests the growing legitimacy of qualitative stories, quotes, and impact statements from participants as evidence in evaluation. For some, this movement away from purely quantitative methods sacrifices the rigor and generalizability of the findings. Others see this change as long-overdue to improve the contextual and multicultural validity of the evaluation findings by considering multiple perspectives and ways of knowing in the evaluation process (Hood, 2004; Mertens & Wilson, 2018).

A culturally responsive approach to evaluation opens the door for evaluators to use innovative methodologies and flexible processes to design and conduct evaluations collaboratively with those whom the program and its evaluation impact. In this paper, we describe a framework for implementing a culturally responsive approach to evaluation in EE that seeks to include stakeholders, particularly those from underrepresented groups, in the evaluation process with specific attention paid to issues of race, power, privilege, and inequity. Stakeholders, also known as rightsholders in indigenous communities (Pomart, 2020), are defined as anyone affected by or with a vested interest in the program, its impact, or the evaluation process (Gold, 1983). In EE, relevant stakeholders could include program staff from various organizational levels, program participants, parents, chaperones, teachers, school administrators, local community members, and program funders. As the field of EE strives to be more relevant to and effective for diverse audiences, taking a culturally responsive approach to evaluation could improve the validity and use of the findings and also create opportunities for the co-creation of programs and increased trust and collaboration between program providers, decision-makers, evaluators, and program participants.

2. Culturally responsive approaches to evaluation

The field of evaluation has identified several approaches that place inclusivity, context, and culture at the center of the evaluation process. These culturally responsive approaches to evaluation have been described in the literature as culturally responsive evaluation (Hood, Hopson, & Kirkhart, 2015), multicultural evaluation (Hopson, 2003), cross-cultural evaluation (Chouinard & Cousins, 2009), transformative evaluation (Mertens, 2008), equity-focused evaluation (Patton, 2012), systems-oriented evaluation (Thomas & Parsons, 2017), collaborative evaluation (Rodriguez-Campos, 2012), participatory evaluation (Whitmore, 1998), democratic evaluation (MacDonald, 1974), and deliberative democratic evaluation (House & Howe, 2000). Authors have begun to distill common principles and practices from these various types of evaluation to put forth a flexible approach to evaluation that considers the needs and perspectives of diverse stakeholders and promotes inclusivity and equity (Askew, Beverly, & Jay, 2012; Boyce & Chouinard, 2017; Chouinard & Cousins, 2009; Greene, 2006; Hood, 2004; Samuels & Ryan, 2011; Thomas & Parsons, 2017). We add to this discussion by first presenting a brief review of the origins of culturally responsive evaluation (CRE), defining CRE, and discussing its key components before detailing a proposed CRE framework with specific attention to its potential application in the context of EE.

2.1. Origins of culturally responsive evaluation

CRE builds off ideas put forth by many prominent evaluation theorists, including Stake's (1975) responsive evaluation, which relied heavily on qualitative methods to understand a program and its impact from the stakeholders' perspectives. Guba and Lincoln (1981) furthered responsive evaluation by combining theory with practice, articulating how responsive evaluation can be used in concert with naturalistic case studies to more holistically understand the multiple realities constructed by stakeholders and evaluators. House (1990) advocated for incorporating social justice into the evaluation process, arguing that all relevant stakeholders were often not included in evaluation processes and that those excluded tended to be in lower socioeconomic classes with less

power. Karen Kirkhart, in her 1994 presidential address at the American Evaluation Association (AEA) conference, proposed the idea of multicultural validity in evaluation. Kirkhart (1995) asserted that the influence of culture should be examined within the contexts of methodological validity, interpersonal validity (concerning the evaluator's personal interactions with subjects), and consequential validity (concerning the changes, intended consequences, and unintended consequences an evaluation imposes on a system). For an evaluation to possess multicultural validity, the methodologies and measures should be relevant to and adequately represent the target population; the evaluator should have a solid understanding of their own culture, values, and norms and how these personal characteristics may impact their communications and interactions with the target population; and the use of the evaluation and its cultural impacts must be examined and determined to be just (Kirkhart, 1995). These ideas inform what is known today as culturally responsive evaluation.

2.2. What is culturally responsive evaluation?

CRE is a collaborative approach to evaluation that intentionally engages relevant stakeholders, particularly those from traditionally underrepresented groups, to design and implement an evaluation process that embraces multiple perspectives and ways of knowing. CRE is sensitive to, reflective of, and valid for the culture and context, and attends to social issues of race, power, privilege, and inequity (Acree & Chouinard, 2020; Askew et al., 2012; Chouinard & Cousins, 2009; Greene, 2006; SenGupta et al., 2004).

Culturally responsive evaluators question normative assumptions based on Western ideals and traditional methodologies that exclude non-dominant epistemologies and fail to consider the personal experiences and histories of program stakeholders (Hall, 2020; SenGupta, Hopson, & Thompson-Robinson, 2004). Evaluators using this approach also resist a deficit mentality that places blame on individuals and cultures for societal problems from the standpoint of the predominant culture (Mayeno, 2000; Tuck, McKenzie, & McCoy, 2014). This approach contests the notion that evaluation should be value-free, conducted by a detached evaluator using standardized measures that consider context and culture as little more than contributing variables (Hood, 2004; Thomas & Parsons, 2017). CRE is fundamentally values-based, accepting the idea that evaluation cannot be separated from the culture and context in which it is located and recognizes that evaluation inherently prioritizes some values over others (Chouinard & Cousins, 2009; Hopson, 2003; Samuels & Ryan, 2011).

Numerous decisions are made throughout the evaluation process, such as identifying relevant evaluation questions and determining appropriate methods of inquiry, data analysis, and interpretation. Many of these decisions are typically made by organizational leaders or evaluation consultants without engaging stakeholders in the decision-making process (Chouinard, 2013). This can perpetuate inequities within the evaluation process by excluding stakeholders' values and opinions. Evaluators using a culturally responsive approach employ transparent and collaborative procedures and decision-making processes to facilitate deliberations among stakeholder groups (Acree & Chouinard, 2020; Thomas & Parsons, 2017). Stakeholders do not need to be involved in every decision; however, they should be involved in establishing the decision-making process, as well as major decisions that build the framework of the evaluation, such as determining the goals, the methods used, and the means by which evaluation results will be communicated.

Engaging stakeholders in collaborative processes, like CRE, heavily relies on building trust between the organization, the evaluator, and the stakeholders (Poth & Shulha, 2008; Taut, 2008). Establishing this trust builds the foundation for meaningful collaboration. If trust is absent among involved parties, CRE will not likely be successful. Stern and colleagues (Stern & Baird, 2015; Stern & Coleman, 2015) identify three forms of trust important in such collaborative processes: rational,

affinitive, and systems-based trust.

Rational trust involves predicting the likely outcomes of the behaviors of one's counterparts in a collaborative process. It is typically based on assessments of competence, consistency, past performance, and goal alignment. Rational trust forms when stakeholders work together effectively and can find ways to demonstrate their competence, when information is openly shared between members, and when group goals are consistently met. Affinitive trust refers to perceptions of positive social relationships between stakeholders. It can emerge through positive social interactions, demonstrations of active listening, and efforts to identify merit in others' statements and express recognition of that merit (Fisher & Shapiro, 2005).

Systems-based trust refers to trust in the processes set up for the group's interactions. Systems-based trust can be fostered when groups work together to outline agreed-upon procedures for effectively collaborating, communicating, and making decisions at the start of the process. Systems that provide a safe space for disagreement and debate tend to yield better decisions (Stern & Predmore, 2011). Systems that create these spaces often involve external facilitation processes that can reveal the underlying interests of the stakeholders involved. Once identified, these interests can help to develop shared criteria by which decisions about program evaluation can be made (Fisher, Ury, & Patton, 2011; Stern & Baird, 2015). For example, if multiple stakeholders can articulate their own desired outcomes of EE programs, these outcomes can then be openly discussed and operationalized within an evaluation. Collaborative forms of evaluation, in which stakeholders play a role in collecting or interpreting data, can further enhance systems-based trust between organizations and their communities, as each can be assured of the veracity of the data (Stern, 2018). Understanding the dynamics of each of these dimensions of trust among stakeholder groups can help evaluators facilitate collaboration and work to build trust in areas where it is lacking.

2.3. Key Components of Culturally Responsive Evaluation

CRE is a collaborative approach that intentionally includes relevant stakeholders, particularly those from traditionally underrepresented groups. CRE emphasizes and promotes collaborative stakeholder engagement throughout the evaluation process. This is particularly important when determining the criteria by which decisions are made such that even if stakeholders are not directly involved in every decision, they feel comfortable with and included in the decision-making process. Stakeholders can also help to identify groups that may be missing from the process as well as help determine the overarching evaluation design including the scope, goals, methodologies, and the means by which the findings will be disseminated (Askew et al., 2012; O'Sullivan, 2004). While collaborative approaches generally aim to include the voices of stakeholders in the evaluation process, CRE is intentional about including stakeholders from marginalized, underrepresented, and underserved groups (Acree & Chouinard, 2020).

Collaborative evaluations are based on the assumption that stakeholders have unique and valuable knowledge about the program, its context, and the audiences served. Taking this one step further, CRE employs a strengths-based approach by recognizing the strengths of stakeholders and communities, as opposed to perpetuating a deficit-based view that is often rooted in perceptions of the dominant culture (Greene, 2006; SenGupta et al., 2004; Thomas & Parsons, 2017). Including stakeholders in the evaluation process has the added benefit of building evaluation capacity and promoting evaluative thinking, or critical thinking applied to the context of evaluation (Buckley, Archibald, Hargraves, & Trochim, 2015). By being involved in the evaluation process, stakeholders are more likely to care about and understand the evaluation findings and put them to use (O'Sullivan, 2012; Torres *et al.* 2000).

Identifying stakeholders to participate in the evaluation process will depend on the context of the organization, the program, and the

evaluation. The challenge is to find a balance between including all or a majority of persons within stakeholder groups, which can make collaboration more complex, and ensuring a representative and heterogeneous group of stakeholders that each uniquely contribute to the process (Luyet, Schlaepfer, Parlange, & Buttler, 2012). Some have proposed using a set of criteria to identify initial stakeholders (Creighton, 1986) followed by a snowball technique where initial stakeholders help to identify additional stakeholders (King, Feltey, & Susel, 2015). Others have identified a more systematic means for identifying stakeholders, including soliciting advice from topical experts and community groups and following that up with an extensive web search to identify additional persons or organizations to invite into the process (Tsang, Barnes, & Dayer, 2021). This type of method, while more time consuming, can help build a more comprehensive and inclusive stakeholder list (Colvin, Witt, & Lacey, 2016). For EE programs, engaging in CRE may mean involving non-traditional stakeholders in the evaluation process, such as school administrators, teachers, students, and parents. Simply engaging diverse stakeholders throughout the process does not constitute CRE, however (Askew et al., 2012). Other key components include considering how culture impacts the evaluation process and attending to issues of power and privilege. Ultimately the principal underpinning stakeholder involvement in CRE is to be more inclusive in the process.

CRE is sensitive to, reflective of, and valid for the culture and context. Culture is a dynamic set of traits, processes, and patterns constructed and shared by a particular group, including preferences, behaviors, life experiences, histories, perceptions, activities, symbols, customs, and institutions (Chouinard & Cousins, 2009; Hopson, 2003). Often, culture can be misinterpreted and oversimplified to mean the physical, outward appearance of a group. This operationalization of culture can be damaging to the evaluation process by relying on stereotypes and faulty assumptions of homogeneity within groups that are defined in this way (Chouinard & Cousins, 2009; Thomas & Parsons, 2017). Culture shapes the worldviews, values, and beliefs of all stakeholders in the evaluation process, including program participants, program staff, and the program evaluator(s). It not only influences assumptions made about programs and their participants, program design and implementation, and participant experiences, but also informs evaluation practices and methodologies (Acree & Chouinard, 2020; Samuels & Ryan, 2011). Culturally responsive evaluators must be reflexive and critically assess how their own assumptions, worldviews, and biases may differ from program stakeholders and influence the evaluation process (Hall, 2020; Hopson, 2003; McBride, 2011; Samuels & Ryan, 2011). To aid in this assessment as time permits, evaluators are encouraged to spend time in the communities served by the programs, develop relationships with program staff and other stakeholders, and observe organizational processes and program implementation to gain better understanding of the multiple perspectives included in the evaluation process (Chouinard & Cousins, 2009).

Context is the social, historical, geographical, economic, and political setting in which the program occurs (Chouinard & Cousins, 2009; Samuels & Ryan, 2011). It can include the sociodemographic characteristics of program participants, the physical setting of the program, organizational structure and leadership, economic conditions of organizations and program participants, program histories, and community histories. Not only is it important for evaluators to examine relevant contextual factors when determining appropriate methods for data collection and analysis, but doing so also provides insights for interpreting the evaluation data. Substantial background research may need to be conducted by evaluators, along with interviews or informal conversations with stakeholders to understand the past, present, and future conditions that influence the program and its implementation. In the case of EE programs, evaluators may need to spend considerable time building relationships with and learning about the communities served by their programs. This could include specific neighborhoods, schools, school districts, agencies, companies, or other organizations.

CRE employs culturally responsive methodologies. CRE does not dictate the use of specific methodologies. It does, however, require that the evaluation methodologies and data collection tools are culturally appropriate for the intended audience (Frierson, Hood, & Hughes, 2002; Kirkhart & Hopson, 2010). This does not necessarily preclude the use of purely quantitative methods if these methods were chosen after considering alternative methods and pilot testing instruments with the target population. However, some scholars have argued that quantitative methods alone do not capture participants' lived experiences articulated by their own voice, which can diminish the multicultural validity of the findings (Stickl Haugen & Chouinard, 2019; Thomas & Parsons, 2017). A literature review of cross-cultural evaluations conducted between 1991 and 2008 revealed that the majority of studies implemented qualitative or mixed-method designs (Chouinard & Cousins, 2009). Specific data collection methods identified in the review included ethnographies, case studies, interviews, focus groups, observations, document analyses, and surveys, as well as more innovative methods including reflexive autoethnographies, storytelling, testimonials, and timelines (Chouinard & Cousins, 2009). This list is not all-encompassing, but it reveals multiple options available for use in CRE and demonstrates that traditional evaluation methods can be culturally responsive if care is taken when designing measurement tools. For example, interviews or open-ended survey items can be used to elicit perceptions of program impact, using rubrics to analyze those data and identify positive and negative impacts as well as specific outcomes achieved. In some cases, these results might be compared to quantitative survey results to build a more complete picture of impacts that emerge from participants' own voices or to clarify trends that emerge from quantitative instruments.

Methodologies used in CRE are often informed by critical race theory and indigenous and decolonizing methods (Boyce & Chouinard, 2017). These methods place the values and ways of knowing of the target population at the forefront of the evaluation and critically examine how societal conditions have perpetuated racism and discrimination (LaFrance & Crazy Bull, 2009; Solorzano, 1997). To develop culturally appropriate methodologies and measures, evaluators should reflect on the histories of marginalized groups and how they have been treated with respect to prior evaluation or research. They should also consult with stakeholders, including program participants, to ensure language, concepts and communication styles reflect local norms; to validate instruments with the target population; and to consider the use of non-traditional methods of data collection (Askew et al., 2012; Chouinard & Cousins, 2009). For example, if a survey or interview guide is used for data collection, members of the target population could examine the instrument and provide suggested edits where the language is not commensurate with their culture, experiences, or comprehension level. After review, the instrument should be tested with the target audience while also providing them with the opportunity to indicate when they do not understand what is being asked and to identify emergent themes or outcomes not initially conceptualized earlier in the process.

When EE organizations conduct evaluations internally, without the help of an experienced evaluator, a common practice is to use pre-existing evaluation tools, such as statistically validated surveys, or to design program-specific surveys to measure knowledge, awareness, and satisfaction. Creating and validating survey instruments can be a long and arduous process, with many iterations before the final product is implemented (Briggs, Trautmann, & Phillips, 2019; Dillman, Smyth, & Christian, 2014). To save EE practitioners valuable time, many EE evaluators and researchers have created standardized surveys that can be applied across a variety of programs. These validated standardized instruments can be found in the literature and many are available on academic and professional organization's websites such as the North American Association of Environmental Education (NAAEE) Evaluation Portal (https://evaluation.naaee.org/).

While standardized instruments can be statistically validated, they may or may not be validated for diverse groups or traditionally

underserved audiences. When using a standardized survey instrument within CRE, the evaluator should analyze the culture and context in which it will be used, collaborate with stakeholders to determine if the survey instrument meets their needs, and consider if other forms of data collection can supplement survey findings. Collecting qualitative data can help evaluators further understand the culture of the organization and communities served to elicit multiple perspectives in a way that reflects the personal experiences and voices of the program participants (Briggs et al., 2019). These data can be used in concert with survey data to gain a more holistic understanding of program impact and the potential explanations as to why certain outcomes are achieved or not achieved. Additionally, modifications could be made to the survey instrument similar to those that are made when translating a survey to another language. In these cases, survey designers must critically examine how intended meanings may shift when translating the document word for word. Specific words present in the original survey may need to be changed to address the same concept (Briggs et al., 2019; Johnson, Pennell, Stoop, & Dorer, 2018). While different cultures may share a common language, variations exist in the communication styles, semantic meanings, and normative uses of the language (Chouinard & Cousins, 2009). Therefore, to adapt a survey for use in a culture that was not originally considered during its design, survey items will likely need to be reworked to improve the construct validity of the measure. This process is best carried out with significant input from program participants and other stakeholders.

CRE attends to social issues of race, power, privilege, and inequity. CRE is firmly situated in the transformative paradigm and social justice branch of evaluation practice as described by Mertens & Wilson (2018). It aims to promote social justice by attending to issues of race, power, privilege, and inequity. By using a collaborative approach and focusing on culture and context, CRE highlights historically marginalized perspectives that enlighten the evaluation process. Evaluators are encouraged to build relationships with, and may choose to act as advocates for, oppressed, underrepresented, and underserved groups to mitigate the imbalances of power and privilege between stakeholder groups (Askew et al., 2012). Privilege is apparent in the field of evaluation, where White, Western epistemologies have been held in higher regard than other ways of knowing (Kirkhart, 2016). Applying privileged ways of knowing to evaluations of programs that serve communities with non-dominant worldviews has severe implications for how constructs are defined and operationalized, and the validity of the evaluation findings (Kirkhart, 2016). A focus on equity also has implications for the data analysis phase of an evaluation, as interpreting impacts using averaged results does little to reveal who is benefiting most and least from the program (Carden, 2017; Greene, 2016). CRE shifts away from

evaluating to understand average program impact to understanding who is impacted, in what ways, and why, by examining the underlying social, political, and historical contexts in which the program and communities served are situated. At each step in the evaluation process, the social issues of race, power, privilege, and inequity are assessed by considering who is included and who is excluded from the evaluation process, whose worldviews and ways of knowing are privileged, and who is benefiting or not from the program.

3. A culturally responsive evaluation framework

We present a culturally responsive evaluation framework adapted from Frierson *et al.* (2002) (Fig. 1) and provide guidance on how to implement this type of evaluation in the field of EE. Using a culturally responsive approach to evaluation adds layers of complexity to an already challenging process. This added complexity will likely be more costly and time-consuming than a traditional evaluation process. While we advise organizations who wish to conduct CRE to advocate for adequate funding from sponsors and to partner with an experienced external evaluator, we acknowledge this is not always feasible. Organizations that conduct evaluations internally without the help of an experienced evaluator can also make use of a culturally responsive approach to evaluation. The degree to which an organization implements a culturally responsive approach to evaluation can vary, but any amount of responsive and reflexive thinking can help an evaluation be more culturally appropriate and valid.

Many of the steps in the framework and critical questions outlined below can also inform the design and implementation of the program itself, not just the evaluation. Ideally, program and evaluation planning occur simultaneously, which improves the likelihood that programs are culturally relevant at the outset and can help link desired programmatic outcomes more closely to the program content and delivery. Both the program and its evaluation have the potential to be culturally responsive and better meet the needs of participants if program stakeholders are engaged in the design of the program and its evaluation. Additionally, by designing and implementing the program and the evaluation in concert, evaluators may be able to identify potential issues as the program is implemented and facilitate opportunities for adaptation to the culture and context (Hall, Freeman, & Roulston, 2014).

At each step in the evaluation process, we propose a set of critical questions for evaluators, program staff, and other stakeholders to consider. The critical questions presented within this framework are not all-encompassing, and some may not be relevant to all program contexts. These questions are meant to serve as a starting point to address the impacts of culture, race, power, privilege, and inequity on the validity of
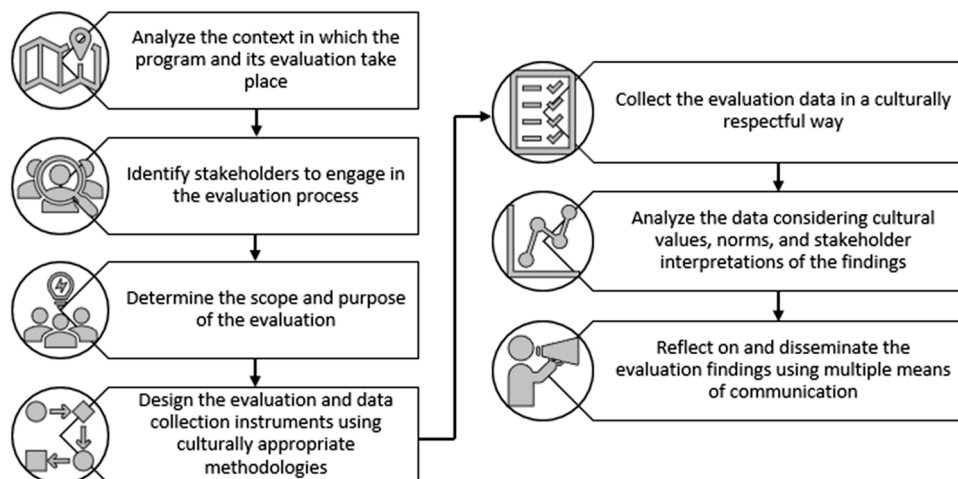

Fig. 1. A culturally responsive evaluation framework.

the evaluation processes and findings, especially in EE contexts. CRE requires reflexivity of the evaluator to constantly re-examine how their values, assumptions, and decisions may impact program stakeholders. Thus, these questions should be asked and answered multiple times throughout an evaluation process. If at any point the evaluator or organization feels the answer to these questions should not be made public or transparent to the stakeholders involved, this may invite further self-reflection and indicates a need to revisit the cultural appropriateness of the program and evaluation design.

3.1. The Evaluation Framework

1. *Analyze the context in which the program and its evaluation take place*

The evaluator should conduct background research on the program and its context to identify and gain a better understanding of the potential contextual factors that may impact the evaluation process (Chouinard & Cousins, 2009). Within EE, these contextual factors may include the extent to which EE is embedded within formal schooling, the accessibility of local green spaces, the locations and severity of undesirable or contaminated land or other environmental issues, the amount of time students typically spend outdoors, access to natural resources and the benefits they provide, the make-up of EE organizational leadership, the values embedded within current EE programming, and where inequities exist. Evaluators are encouraged to practice reflexive thinking to identify areas in which the culture and context diverge from their own experiences.

Critical questions:

- What are the social, political, historical, and environmental conditions and context of the program?
- If an existing program, who was involved in the planning and design? What was the process for designing the program? Who decided what the intended outcomes should be?
- What assumptions are being made related to the program design, implementation, and its intended impacts?
- What are the perceived costs and benefits created by the program? How are they distributed across stakeholder groups?
- What worldviews pertaining to the environment were considered when designing this program? How might these worldviews impact participants, especially those who do not share the worldviews?
- Where is the program physically located and how might its location act as a barrier to underserved audiences?
- How might historical and environmental racism and discrimination impact who participates and the outcomes achieved by underserved audiences?
- What cultures are represented in the program staff, stakeholders, and evaluators? In what ways do these cultures overlap or diverge from each other?
- Do program staff represent participants in terms of racial, ethnic, and cultural characteristics? Are marginalized and underserved groups adequately represented by program staff?
- Who are the program participants? Are certain groups absent or underrepresented among program participants?
- Do all stakeholders commit to conducting collaborative evaluation processes using a culturally responsive approach?
- Are there adequate resources available to conduct a culturally responsive evaluation?
- Do funder expectations align with a culturally responsive approach to evaluation?

2. *Identify stakeholders to engage in the evaluation process*

Before designing the evaluation, the evaluator should conduct a careful analysis of who should be included in the evaluation process,

with specific attention paid to including stakeholders from marginalized and traditionally underrepresented groups. CRE is fundamentally a collaborative process, engaging stakeholders in all aspects of the evaluation. By ensuring that all relevant stakeholder groups are represented in the evaluation process, the evaluation is more likely to possess multicultural validity by addressing relevant questions, using culturally appropriate methods, and interpreting the data with respect to culture and context (Askew et al., 2012; Brandon, 1998). In the context of EE, relevant stakeholders could include program participants, students and their parents, teachers and school administrators, program funders, and program staff from various organizational levels; although, a full stakeholder analysis may be needed to identify groups that are often left out of the initial search (Tsang *et al.* 2021). The threshold for adequate representation can vary by context, and there will always be trade-offs between cost, time, and full inclusion of all stakeholders and stakeholder groups. The intent is to ensure full representation of the breadth of cultures, identities, and perspectives in the evaluation process. Note, too, that compensation or other meaningful recognition for stakeholders for their role in the evaluation should be considered, especially when they have previously been subjected to extractive or otherwise oppressive research or evaluation experiences. Compensation and recognition can take many forms, such as providing stipends, transportation, or meals; providing co-authorship on published reports; or highlighting stakeholders' efforts within organizational communication (e.g., social media posts).

Critical questions:

- Who is currently involved in the evaluation process and to what extent?
- Are marginalized, underserved, and underrepresented groups adequately represented? Note: A single person from a group or culture may not represent the diversity of perspectives of that group so it is important to seek multiple perspectives within stakeholder groups.
- Who is missing from the evaluation process?
- How can stakeholder groups who are not currently or traditionally involved in the evaluation process be meaningfully engaged?
- Are there imbalances of power and privilege among stakeholder groups?
- How might teachers, school administrators, parents, and program participants be engaged in the evaluation process, acknowledging competing responsibilities and the limited time these groups have to contribute?
- Are there specific populations that are currently not being served by the program that should be included in the evaluation process?

3. *Determine the scope and purpose of the evaluation*

CRE relies on collaborative processes to facilitate discussions among stakeholder groups and between these groups and the evaluator to determine the scope and purpose of the evaluation as well as define the guiding evaluation questions. This may often require time spent defining key terminology used in evaluation so that all stakeholders can understand each other. A decision-making process and the criteria by which decisions will be made should be discussed and agreed-upon by participating stakeholders. This process should make clear how stakeholders' interests will be considered and incorporated even if they are not present when the decisions are made. Revisiting the core motivations and definitions of key terms as decisions are made throughout the evaluation process can help guide evaluators toward just and equitable designs. In the context of EE, many organizations share a distinct mission—to build an environmentally literate public that has the knowledge, skills, dispositions, and motivations to address pressing and complex environmental issues (see UNESCO, 1977). Seeking areas of alignment between the mission-based goals of the organization and stakeholder interests can chart a path toward mutually beneficial

evaluation outcomes.

Critical questions:

- Does the evaluation process foster the building of trust between stakeholders, program staff, and evaluators?
  o In particular, consider elements of rational, affinitive, and systems-based trust, discussed earlier.
- Can stakeholders agree on a process for decision-making that feels fair and safe?
- How might imbalances of power and privilege impact the decision-making processes?
- Have all relevant stakeholder interests been identified and considered?
- Can stakeholders agree upon shared interests in the program? Can those interests be used to develop criteria for evaluating potential program design or evaluation elements?
- Does the program logic align with proposed evaluation questions?
- Are stakeholders who are not traditionally involved in the evaluation process, including teachers, school administrators, parents, program participants, and other relevant community members empowered to participate in the process?
- Are program participants encouraged to speak up and advocate for their interests within the program planning and evaluation process? How might the evaluator mitigate power imbalances stemming from age differences?
- Are program funders, directors, and managers willing to embrace multiple perspectives in the decision-making process when defining evaluation questions?

4. *Design the evaluation and data collection instruments using culturally appropriate methodologies*

Once the scope, purpose, and evaluation questions are determined, the evaluation can be designed to meet those objectives. Program staff are likely best positioned to identify feasibility concerns with regard to the evaluation design, while program participants and other stakeholders can provide input on how different methodologies might align with or contradict cultural traditions and norms. If a sampling scheme will be used, evaluators are encouraged to choose a sample that ensures all participant groups are equitably represented. Data collection processes should be developed specifically for the target audiences with stakeholder input to ensure they are culturally commensurate. Specific measures, whether new or pre-existing, should align with the evaluation questions developed collaboratively by all relevant stakeholders. Pilot testing evaluation methods and instruments with target audiences can identify issues and unintended meanings that need to be rectified before implementation.

Critical questions:

- What are all the various ways in which the evaluation questions can be addressed, considering quantitative, qualitative, and mixed methodologies?
- Which of the identified designs and methodologies most closely align with cultural norms?
- Does the sampling scheme ensure representation of marginalized or underrepresented groups? This may require over-sampling of particular populations to ensure evaluation questions are answered adequately.
- Do the data collection instruments conform to the language and communication styles of program participants?
- In what ways can the data collection instruments be validated for the target audience?
- Does pilot testing reveal any potential issues with the evaluation instrument?

- Are the data collection instruments at an appropriate reading and comprehension level for the age of program participants?
- How can program providers, teachers, students, parents, and other stakeholders inform data collection procedures and instruments that are compatible with program participants?

5. *Collect the evaluation data in a culturally respectful way*

After the evaluation and associated measures are designed and pilot tested, the evaluation can be implemented. The evaluator should consider communities' past and current experiences and relationships with external researchers, evaluators, and others in a position of power and seek to acknowledge and directly address negative perceptions of research and evaluation. This entails upfront, honest communication about the organizational goals and intentions undergirding the work. Evaluators are encouraged to shift their perspectives from conducting evaluations *on* program participants to conducting evaluations *with* program participants. This shift in thinking can create a more welcoming atmosphere, empower program participants, and result in more valid data. As previously mentioned, compensation for participants' time may be important in some circumstances.

Critical questions:

- Who is providing the evaluation data?
- How might the complex history of research with marginalized groups influence participants' willingness to cooperate and engage in the process?
- How can evaluators communicate the value of the evaluation to participants in a culturally meaningful way?
- Have evaluators listened sufficiently to understand stakeholders' interests and concerns, addressed them in their evaluation design, and communicated how they have been addressed?
- How can participants be ensured of safety from any potential perception of harm? Who should communicate with them about this?
- Are those responsible for collecting data knowledgeable of the cultural context?
- How might the prevalence of standardized tests influence how participants/students perceive data collection methods such as surveys?
- How might the presence of peers and adults influence student behaviors and self-reports? How will this impact the evaluation findings?

6. *Analyze the data considering cultural values, norms, and stakeholder interpretations of the findings*

The evaluator should analyze the data using an iterative process, frequently seeking stakeholder input on the interpretation of the findings to ensure cultural validity. Collaborative and shared analyses, using approaches like 'data parities,' which include stakeholders in the analysis process (Westaby, Williams, Robinson, & Connors, 2019), may also be considered. When appropriate, data should be disaggregated to understand differences and similarities among groups and the diversity present within them.

Critical questions:

- What contextual factors need to be considered when interpreting the data?
- How can different forms of data help to tell a cohesive story?
- What convergent and divergent interpretations of the data are present among stakeholders?
- Are stakeholder interpretations of the data reflected in the evaluation findings?

- How do the interpretations of different stakeholders align with or diverge from each other? What cultural or contextual factors might explain these differences?
- Which cultural groups are included in the data and how can they be aggregated and disaggregated to understand the nuances between and within groups?

7. *Reflect on and disseminate the evaluation findings using multiple means of communication*

Upon sharing evaluation findings widely, particularly to stakeholder groups that were engaged in the evaluation process, the evaluator, along with colleagues and collaborators, should reflect on the implications of the evaluation findings paying attention to whose voices are present within the data and how various groups are benefiting or not from the program. Often, a formal evaluation report is required by funders. However, this report is not the only way in which evaluation findings should be disseminated. Different stakeholder groups may have various understandings of traditional research and evaluation nomenclature. Evaluators are encouraged to find innovative ways to disseminate evaluation findings, such as interactive presentations, narratives, or skits (Johnson, Hall, Greene, & Ahn, 2013). Program stakeholders may also be encouraged to be involved in presenting or otherwise disseminating the evaluation findings.

Critical questions:

- Are certain groups over or underrepresented within the data? How might this impact the interpretation of the findings?
- Do certain groups disproportionately benefit from the program?
- What are the primary interests of different audiences in the evaluation results?
- How can stakeholders be involved in disseminating the evaluation findings?
- How can the evaluation findings best be presented to encourage their use?
- What means of communication are best suited for different stakeholder groups?
- Are the findings presented in a way that respects the cultural context?

4. Lessons learned: a call to action

At a time when the field of EE is in need of more rigorous evaluation methods and looking to make actionable steps towards being more diverse and inclusive, CRE can help to achieve both of these objectives. Engaging in the process of CRE can uncover implicit biases and assumptions, highlight and rectify a lack of diversity within organizational leadership and decision-making processes, and improve the accessibility and impact of programs for diverse and underserved audiences.

Fully operationalizing this framework is resource intensive and can require high levels of capacity to facilitate. Not all organizations will have the finances, time, stakeholder relations, and staff capacity to fully implement CRE. CRE may thus be easier to conceptualize for specific programs, as opposed to broader evaluation systems of the full suite of offerings provided by an EE organization. Specific programs and their evaluations can be co-created (designed in partnership) with representatives of the target audience(s). In cases where stakeholder engagement may be lacking, organizational leadership is encouraged to approach the guiding questions from multiple perspectives and practice reflexivity.

This framework is aspirational, and the level at which stakeholders can or want to engage in the process will vary by context. While the main thrust of CRE is inclusiveness, endless stakeholder engagement with all relevant parties is of course unrealistic. Thus, decisions need to be made about the key elements of program and evaluation planning and implementation that call for broad engagement. The distinction between developing criteria for how to make decisions and actual decision-making (Fisher et al., 1991) may represent a reasonable threshold. While close partners may be involved in actual decision-making, decision-making about program or evaluation specifics in large groups may be untenable. If collaborative efforts are undertaken to identify the key interests of all relevant stakeholders, focusing on ensuring decision-making processes are transparent (i.e., that efforts are taken and communicated to develop strategies that address those interests) may be sufficient and more realistic.

While we encourage staff members, directors, and managers internal to an EE organization to make use of this proposed framework and critical guiding questions, the overall approach is intended to be a participatory process involving external stakeholders. External evaluation experts can be instrumental in these processes by serving both as methodological experts and as facilitators, paying careful attention to the stakeholder engagement process throughout. We urge funding agencies that work with EE organizations to recognize the importance of the CRE approach and to provide adequate funding and lenient timelines for organizations to properly evaluate their programs in a way that is sensitive to, reflective of, and valid for the culture and context in which the program is located. This may include additional funds to support stakeholders as collaborators in the process commensurate with the time and expertise they provide. We also urge EE organizations to advocate for the resources needed to conduct such evaluations. Finally, we ask organizations and evaluators to invite researchers to examine their process of CRE as case studies. What lessons can we learn? What seems to work for whom and in what contexts? Building a body literature of these cases can enhance the basic framework we present here and keep these ideas at the forefront of the field.

5. Conclusion

CRE is fundamentally a collaborative approach and can serve to build evaluation capacity and promote evaluative thinking within organizations and across communities. Using this approach to evaluation not only supports organizational learning, but also furthers social justice efforts by asking critical questions and produces evaluation findings that are culturally valid. The critical questions posed here will help evaluators and educators practice reflexive thinking, and many can be used throughout the program design process as well as the evaluation itself, which we argue should be intertwined. Using a culturally responsive approach to evaluation can also create opportunities for the co-creation of programs and increased trust and collaboration between program providers, evaluators, and program participants. Designing a program and its evaluation in collaboration with the communities served can push the field towards being more culturally responsive in all that we do, making EE more accessible, meaningful, and relevant to all communities.

CRediT authorship contribution statement

Kelley C. Anderson: Conceptualization, Writing – original draft, Writing – review & editing **Marc J. Stern**: Conceptualization, Writing – review & editing, Supervision. **Robert B. Powell**: Conceptualization, Writing – review & editing. **Ashley A. Dayer**: Conceptualization, Writing – review & editing. **Thomas G. Archibald**: Conceptualization, Writing – review & editing.

Declaration of Interest

none.

References

Acree, J., & Chouinard, J. A. (2020). Exploring use and influence in culturally responsive approaches to evaluation: a review of the empirical literature. *American Journal of Evaluation, 41*(2), 201–215.

Ardoin, N. M., Bowers, A. W., Roth, N. W., & Holthuis, N. (2018). Environmental education and K-12 student outcomes: a review and analysis of research. *The Journal of Environmental Education, 49*(1), 1–17.

Ardoin, N. M., Clark, C., & Kelsey, E. (2013). An exploration of future trends in environmental education research. *Environmental Education Research, 19*(4), 499–520.

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

Bourke, N., Buskist, C., & Herron, J. (2014). Residential environmental education center program evaluation: an ongoing challenge. *Applied Environmental Education & Communication, 13*(2), 83–90.

Boyce, A. S., & Chouinard, J. A. (2017). Moving beyond the buzzword: a framework for teaching culturally responsive approaches to evaluation. *Canadian Journal of Program Evaluation, 32*(2), 266–279.

Brandon, P. R. (1998). Stakeholder participation for the purpose of helping ensure evaluation validity: Bridging the gap between collaborative and non-collaborative evaluations. *American Journal of Evaluation, 19*(3), 325–337.

Briggs, L., Trautmann, N., & Phillips, T. (2019). Exploring challenges and lessons learned in cross-cultural environmental education research. *Evaluation and Program Planning, 73*, 156–162.

Buckley, J., Archibald, T., Hargraves, M., & Trochim, W. M. (2015). Defining and teaching evaluative thinking: Insights from research on critical thinking. *American Journal of Evaluation, 36*(3), 375–388.

Carden, F. (2017). Building evaluation capacity to address problems of equity. *New Directions for Evaluation, 2017*(154), 115–125.

Carleton-Hug, A., & Hug, J. W. (2010). Challenges and opportunities for evaluating environmental education programs. *Evaluation and Program Planning, 33*(2), 159–164.

Carman, J. G., & Fredericks, K. A. (2008). Nonprofits and evaluation: empirical evidence from the field. *New Directions for Evaluation, 2008*(119), 51–71.

Chouinard, J. A. (2013). The case for participatory evaluation in an era of accountability. *American Journal of Evaluation, 34*(2), 237–253.

Chouinard, J. A., & Cousins, J. B. (2009). A review and synthesis of current research on cross-cultural evaluation. *American Journal of Evaluation, 30*(4), 457–494.

Colvin, R. M., Witt, G. B., & Lacey, J. (2016). Approaches to identifying stakeholders in environmental management: Insights from practitioners to go beyond the 'usual suspects'. *Land Use Policy, 52*, 266–276.

Creighton, J. (1986). *Managing conflicts in public involvement settings: training manual for Bonneville Power Administration*. Palo Alto, CA: Creighton and Creighton.

Crohn, K., & Birnbaum, M. (2010). Environmental education evaluation: time to reflect, time for change. *Evaluation and Program Planning, 33*(2), 155–158.

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, phone, mail, and mixed-mode surveys: the tailored design method*. Hoboken, NJ: John Wiley & Sons,.

Finney, C. (2014). *Black faces, white spaces: Reimagining the relationship of African Americans to the great outdoors*. Chapel Hill: The University of North Carolina Press.

Fisher, R., & Shapiro, D. (2005). *Beyond reason: Using emotions as you negotiate*. New York, NY: Penguin.

Fisher, R., Ury, W. L., & Patton, B. (2011). *Getting to Yes: Negotiating Agreement Without Giving in* (2nd ed..,). New York, NY: Penguin,.

Frierson, H. T., Hood, S., & Hughes, G. (2002). Strategies that address culturally responsive evaluation. In J. Frechtling (Ed.), *The 2002 User-Friendly Handbook for Project Evaluation* (pp. 63–73). Arlington, VA: National Science Foundation.

Gold, N. (1983). StakeHolder And Program Evaluation: Characterizations And Reflections. *New Directions for Program Evaluation, 17*, 63–72.

Greene, J. C. (2010). Serving the public good. *Evaluation and Program Planning, 33*(2), 197–200.

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

Greene, J. C. (2016). Advancing equity: Cultivating an evaluation habit. In S. I. Donaldson, & R. Picciotto (Eds.), *Evaluation for an Equitable Society* (pp. 49–66). Charlotte, NC: Information Age Publishing.

Guba, E. G., & Lincoln, Y. S. (1981). Effective evaluation: Improving the usefulness of evaluation results through responsive and naturalistic approaches. San Francisco. *CA*. Jossey-Bass,.

Hall, J., Freeman, M., & Roulston, K. (2014). Right timing in formative program evaluation. *Evaluation and Program Planning, 45*, 151–156.

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.

Hood, S. (2004). A journey to understand the role of culture in program evaluation: Snapshots and personal reflections of one African American evaluator. *New Directions for Evaluation, 2004*(102), 21–37.

Hood, S., Hopson, R. K., & Kirkhart, K. E. (2015). Culturally responsive evaluation. In K. E. Newcomber, H. P. Hatry, & J. S. Wholey (Eds.), *Handbook of Practical Program Evaluation* (4th ed..,, pp. 281–318). Hoboken, NJ: Wiley.

Hoole, E., & Patterson, T. E. (2008). Voices from the field: evaluation as part of a learning culture. *New Directions for Evaluation, 119*, 93–113.

Hopson, R. (2003). Overview of multicultural and culturally competent program evaluation. Oakland, CA: Social Policy Research Associates.

House, E. R. (1990). Methodology and justice. *New Directions for Program Evaluation, 1990*(45), 23–36.

House, E. R., & Howe, K. R. (2000). Deliberative democratic evaluation in practice. In D. L. Stufflebeam, G. F. Madaus, & T. Kellaghan (Eds.), , *Evaluation models* (pp. 409–421). Boston, MA: Kluwer.

Johnson, J., Hall, J., Greene, J. C., & Ahn, J. (2013). Exploring alternative approaches for presenting evaluation results. *American Journal of Evaluation, 34*(4), 486–503.

Johnson, T. P., Pennell, B. E., Stoop, I. A., & Dorer, B. (Eds.). (2018). *Advances in comparative survey methods: Multinational, multiregional, and multicultural contexts*. New York, NY: John Wiley & Sons.

Keene, M., & Blumstein, D. T. (2010). Environmental education: a time of change, a time for change. *Evaluation and Program Planning, 33*(2), 201–204.

King, C. S., Feltey, K. M., & Susel, B. O. N. (2015). The question of participation: Toward authentic public participation in public administration. In N. C. Roberts (Ed.), *The age of direct citizen participation* (pp. 391–408). New York: Routledge.

Kirkhart, K.E. (1995). 1994 conference theme: Evaluation and social justice seeking multicultural validity: A postcard from the road. *Evaluation Practice, 16*(1), 1–12.

Kirkhart, K.E., & Hopson, R. (2010). *Strengthening evaluation through cultural relevance and cultural competence*. Paper presented the American Evaluation Association/Centers for Disease Control 2010 Summer Evaluation Institute.

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

LaFrance, J., & Crazy Bull, C. (2009). Researching ourselves back to life: Taking control of the research agenda in Indian country. In D. M. Mertens, & P. E. Ginsburg (Eds.), *Handbook of Social Research Ethics* (pp. 135–149). Thousand Oaks, CA: Sage.

Leeming, F. C., Dwyer, W. O., Porter, B. E., & Cobern, M. K. (1993). Outcome research in environmental education: a critical review. *The Journal of Environmental Education, 24*(4), 8–21.

Luyet, V., Schlaepfer, R., Parlange, M. B., & Buttler, A. (2012). A framework to implement stakeholder participation in environmental projects. *Journal of environmental management, 111*, 213–219.

MacDonald, B. (1974). Evaluation and the control of education. In D. A. Tawney (Ed.), *Curriculum Evaluation Today: Trends and Implications* (pp. 125–136). London, England: Macmillan.

Mayeno, A.S. (2000). Environmental Education Needs and Preferences of an Inner City Community of Color (Master thesis). San Francisco State University.

McBride, D. F. (2011). Sociocultural theory: Providing more structure to culturally responsive evaluation. *New Direction for Evaluation, 131*, 7–13.

McLean, S. (2013). The whiteness of green: Racialization and environmental education. *The Canadian Geographer/Le Géographe Canadien, 57*(3), 354–362.

Mertens, D. M. (2008). *Transformative research and evaluation*. New York, NY: Guilford Publications.

Mertens, D. M., & Wilson, A. T. (2018). *Program evaluation theory and practice*. New York, NY: Guilford Publications.

Monroe, M. C. (2010). Challenges for environmental education evaluation. *Evaluation and Program Planning, 33*(2), 194–196.

O'Sullivan, R. G. (2004). *Practicing evaluation: a collaborative approach*. Thousand Oaks, CA: Sage.

O'Sullivan, R. G. (2012). Collaborative evaluation within a framework of stakeholder-oriented evaluation approaches. *Evaluation and Program Planning, 35*(4), 518–522.

Patton, M. Q. (2002). *Qualitative Research and Evaluation Methods*. Thousand Oaks. CA: Sage.

Patton, M. Q. (2012). Developmental evaluation for equity-focused evaluations. In M. Bamberger, & M. Segone (Eds.), *Evaluation for equitable development results* (pp. 102–114). New York, NY: UNICEF.

Patton, M. Q. (2003). Utilization-focused evaluation. In T. Kellaghan, & D. L. Stufflemean (Eds.), *International handbook of educational evaluation* (pp. 223–242). Netherlands: Springer.

Pease, J. L. (2015). Parks and underserved audiences: an annotated literature review. *Journal of Interpretation Research, 20*(1), 11–56.

Pomart, P. N. (2020). Reframing Indigenous Peoples from Stakeholders to Rightsholders. In S. Taneja (Ed.), *Academy of Management Proceedings* (p. 20874). Briarcliff Manor, NY: Academy of Management.

Poth, C. A., & Shulha, L. (2008). Encouraging stakeholder engagement: a case study of evaluator behavior. *Studies in Educational Evaluation, 34*(4), 218–223.

Powell, R. B., Stern, M. J., & Ardoin, N. (2006). A sustainable evluation framework and its application. *Applied Environmental Education and Communication, 5*(4), 231–241.

Rickinson, M. (2001). Learners and learning in environmental education: a critical review of the evidence. *Environmental Education Research, 7*(3), 207–320.

Roberts, N.S. (2007). *Visitor/non-visitor use constraints: exploring ethnic minority experiences and perspectives*. Gen. Tech. Rep. Golden Gate National Recreation Area, National Park Service. San Francisco, CA: San Francisco State University.

Rodriguez-Campos, L. (2012). Advances in collaborative evaluation. *Evaluation and program planning, 35*(4), 523–528.

Samuels, M., & Ryan, K. (2011). Grounding evaluations in culture. *American Journal of Evaluation, 32*(2), 183–198.

SenGupta, S., Hopson, R., & Thompson-Robinson, M. (2004). Cultural competence in evaluation: an overview. *New Directions for Evaluation, 2004*(102), 5–19.

Solorzano, D. G. (1997). Images and words that wound: critical race theory, racial stereotyping, and teacher education. *Teacher Education Quarterly, 24*(3), 5–19.

Stake, R. E. (1967). The countenance of educational evaluation. *Teachers College Record, 68*, 523–540.

Stake, R.E. (1975). Program Evaluation, Particularly Responsive Evaluation. Kalamazoo, MI: Western Michigan University Evaluation Center.

Stern, M. J. (2018). *Social science theory for environmental sustainability: A practical guide.* Oxford University Press.

Stern, M. J., & Baird, T. D. (2015). Trust ecology and the resilience of natural resource management institutions. *Ecology and Society, 20*(2), 14.

Stern, M. J., & Coleman, K. J. (2015). The multidimensionality of trust: applications in collaborative natural resource management. *Society & Natural Resources, 28*(2), 117–132.

Stern, M. J., Powell, R. B., & Hill, D. (2014). Environmental education program evaluation in the new millennium: what do we measure and what have we learned? *Environmental Education Research, 20*(5), 581–611.

Stern, M. J., & Predmore, S. A. (2011). Decision making, procedural compliance, and outcomes definition in US forest service planning processes. *Environmental Impact Assessment Review, 31*(3), 271–278.

Stickl Haugen, J., & Chouinard, J. A. (2019). Transparent, translucent, opaque: exploring the dimensions of power in culturally responsive evaluation contexts. *American Journal of Evaluation, 40*(3), 376–394.

Taut, S. (2008). What have we learned about stakeholder involvement in program evaluation? *Studies in Educational Evaluation, 34*(4), 224–230.

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

Torres, R. T., Stone, S. P., Butkus, D. L., Hook, B. B., Casey, J., & Arens, S. A. (2000). Dialogue and reflection in a collaborative evaluation: stakeholder and evaluator voices. *New Directions for Evaluation, 85*, 27–38.

Tsang, E. M., Barnes, J. C., & Dayer, A. A. (2021). A web-based approach to stakeholder analysis for identifying and understanding broader constituencies in wildlife conservation. *Society & Natural Resources, 34*(8), 1133–1146.

Tuck, E., McKenzie, M., & McCoy, K. (2014). Land education: Indigenous, post-colonial, and decolonizing perspectives on place and environmental education research. *Environmental Education Research, 20*(1), 1–23.

UNESCO, U. (1977). The Tbilisi Declaration. In Intergovernmental Conference on Environmental Education (pp. 14–26). Tbilisi, USSR: UNESCO.

Warren, K., Roberts, N. S., Breunig, M., & Alvarez, M. A. T. G. (2014). Social justice in outdoor experiential education: a state of knowledge review. *Journal of Experiential Education, 37*(1), 89–103.

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, WI: ¡Milwaukee Evaluation!, Inc.

Whitmore, E. (1998). Understanding and Practicing Participatory Evaluation. *New directions for evaluation, 80*, 1–104.

Wojcik, D. J., Biedenweg, K., McConnell, L., & Iyer, G. (2013). Current trends in environmental education. *Across the Spectrum*, 48–63.

Kelley C. Anderson is a Postdoctoral Research Associate in the Department of Forest Resources and Environmental Conservation at Virginia Tech, where she earned her Ph.D. in 2021. Her research focuses on environmental education, adaptive management, evaluation, and diversity, equity, and inclusion. In recognition of her work and service at Virginia Tech, she received the 2021 A.B. Massey Outstanding Doctoral Student Award for the Department of Forest Resources and Environmental Conservation.

Marc J. Stern is a Professor in the Department of Forest Resources and Environmental Conservation at Virginia Tech, where his teaching and scholarship focus on environmental education, interpretation, and communication; evaluation research; and the human dimensions of natural resource management and sustainability. His recent book, Social Science Theory for Environmental Sustainability: A Practical Guide, published by Oxford University Press, serves as a field guide for conservation practitioners and graduate students to translate social science knowledge into practical application.

Robert B. Powell is the George B. Hartzog, Jr. Endowed Professor in the Department of Parks, Recreation, and Tourism Management and the Dept. of Forestry and Environmental Conservation at Clemson University and the Director of the Institute for Parks, which is an interdisciplinary institute focused on providing research, training, and outreach to support park and protected area management. His research and outreach program focuses on environmental education, interpretation, park and protected areas, and sustainable tourism and he is the co-editor of the Journal of Interpretation Research.

Ashley Dayer is an Assistant Professor of Human Dimensions in the Department of Fish and Wildlife Conservation at Virginia Tech. Her applied conservation research aims to understand what drives people's natural resources-related behaviors. Her research is well-known for informing agencies' and organizations' policies and programs due to her investment in working closely with those who will be applying research. Dr. Dayer holds a Ph.D. in Natural Resources from Cornell University, a M.S. in Human Dimensions of Natural Resources from Colorado State University, and a B.A. in Environmental Science and Public Policy from Harvard University.

Thomas Archibald is an Associate Professor and Extension Specialist in the Department of Agricultural, Leadership, and Community Education at Virginia Tech, where he also directs the Feed the Future Senegal Youth in Agriculture project. A winner of the American Evaluation Association (AEA) Marcia Guttentag Promising New Evaluator Award, he serves on the Board of Directors of the Eastern Evaluation Research Society and is an Associate Editor of Evaluation and Program Planning. He received his PhD in Adult and Extension Education from Cornell University in 2013.