Investigating aspects of diversity, equity, and inclusion, adaptive management, and evaluation in environmental education

Kelley C. Anderson

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
In
Forest Resources and Environmental Conservation

Marc J. Stern
Robert B. Powell
Ashley A. Dayer
Thomas G. Archibald

April 26, 2021
Blacksburg, VA

Keywords: environmental education, diversity, equity, inclusion, adaptive management, evaluation, cultural responsiveness

Investigating aspects of diversity, equity, and inclusion, adaptive management, and evaluation in environmental education

Kelley C. Anderson

Abstract

This dissertation investigates aspects of diversity, equity, and inclusion, adaptive management, and evaluation in environmental education (EE), and is composed of and introduction, three stand-alone manuscripts (Chapters 2-4), and a conclusion. The introduction gives a brief overview of EE and explains why the topics of study included in this dissertation are important to address. In Chapter 2, we used pre-experience student surveys to understand how preparation, adult support, and students' racial identities influence student attendance to a residential environmental education program. Chapter 3 identifies areas and approaches for improving evaluation processes in EE and practitioner satisfaction with those processes. We collected these data using an online survey promoted on social media and emailed to EE practitioners involved in the North American Association for Environmental Education and the Association of Nature Center Administrators. Chapter 4 outlines a culturally responsive evaluation framework for use in EE. The results of these studies show there is far more to be done in the field of EE to create a welcoming and inclusive space for all audiences, to promote the use of evaluation as a tool for continuous learning and improvement, and to ensure evaluations are valid for and reflective of the culture of program participants. The conclusion discusses the two prevalent themes embedded in these manuscripts, namely diversity, equity and inclusion, and evaluation in EE, and ends with a reflection on my time here as a Ph.D. student and where I see my career path heading. This dissertation is meant to provide ideas and suggestions to environmental education practitioners that they can implement in hopes of improving EE and evaluation to meet the needs of all audiences and to address global environmental challenges.

Investigating aspects of diversity, equity, and inclusion, adaptive management, and evaluation in environmental education

Kelley C. Anderson

General Audience Abstract

To work collaboratively and find solutions to the sustainability challenges and social justice issues we face as a society requires new generations to obtain a wide range of knowledge, skills, and motivations. Environmental education (EE) has been shown to equip students with these necessary skills, including increasing knowledge and awareness about environmental issues, enhancing or changing attitudes as they relate to the environment, promoting environmentally responsible behaviors, and building critical thinking, leadership, and collaboration skills. There is a growing recognition that the field of EE must continually adjust and improve its programming to meet the needs of all audiences and to remain relevant for our ever-changing world. The foundations of EE stem from White, Eurocentric values and world views, which can perpetuate inequities in program attendance, participation, and impact between Black, Indigenous, and People of Color (BIPOC) and White people. The research included in this dissertation investigates aspects of diversity, equity, and inclusion, adaptive management, and evaluation in EE. Results show there is far more to be done in the field of EE to create a welcoming and inclusive space for all audiences, to promote the use of evaluation as a tool for continuous learning and improvement, and to ensure evaluations are valid for and reflective of the culture of program participants. The conclusion of this dissertation discusses two prevalent themes embedded in these manuscripts, namely diversity, equity and inclusion, and evaluation in EE, and ends with a reflection on my time here as a Ph.D. student and where I see my career path heading. This dissertation is meant to provide ideas and suggestions to environmental education practitioners that they can implement in hopes of improving EE and evaluation to meet the needs of all audiences and to address global environmental and social challenges.

Acknowledgements

I am so fortunate to have a great support system that helped me thrive and succeed during my time here at Virginia Tech. While there were many involved in this process, I'd like to thank those who were integral to my success at completing my doctorate degree.

To my advisor, Marc Stern, I am thankful every day that you chose to bring me on as your student and guide me throughout the process. Your guidance, insights, and support helped me grow intellectually and personally. You improved my self-confidence and helped shape my identity as a researcher. I look forward to collaborating for many years to come.

To my committee, Bob Powell, Ashley Dayer, and Tom Archibald, thank you for always making time for me and supporting me through this process. Your feedback and guidance gave me new perspectives and expanded my understanding of this field and academia in general. A special thanks to Ashley Dayer and the Dayer lab for inviting me to participate in meetings with your lab group. I learned so much from the Dayer lab and made invaluable connections with those in other departments.

Thank you to my lab mates, Jennifer Brousseau, Caleb O'Brien, Emily Thorpe, and Malia Pownall for making this experience an enjoyable one. Your support, feedback, and friendship mean the world to me, and I am lucky to have had the opportunity to work alongside you for the betterment of the environment and social justice. Another big thank you to Joseph Osborne and Joe Kozak for giving me inspiration and support throughout this process.

Thank you to the NorthBay Education Foundation and NorthBay staff for supporting and collaborating on this work, and to the students and classroom teachers who participated in our study. Thank you to the Pieces Foundation for funding part of this work and to the North American Association for Environmental Education (NAAEE), the Association of Nature Center Administrators (ANCA) for helping to disseminate practitioner surveys to those across the United States and abroad.

Thank you to the Forest Resources and Environmental Conservation Department at Virginia Tech for supporting me throughout my degree and providing funding when times were uncertain during the COVID-19 pandemic.

Thank you to my parents and sisters for always encouraging me to follow my passions and for instilling and shaping my morals, ethics, and integrity that has made me who I am today.

Table of Contents

List of Tables

List of Figures

Attribution

While the majority of the work included in this dissertation was primarily conducted by me, it was a collaborative effort. Chapters 2-4 are standalone manuscripts intended for publication. These papers are co-authored by my committee, Marc Stern, Bob Powell, Ashley Dayer, and Tom Archibald. These co-authors contributed to these manuscripts by suggesting specific analyses, improving the writing, and helping with the organization.

CHAPTER 1

Introduction

Proactive steps can be taken to equip new generations with the knowledge, skills, and motivations necessary to tackle the critical sustainability challenges and social justice issues we face as a society. I refer to both sustainability and social justice because they are inextricably linked (Balaceanu, Apostol, & Penu, 2012; Sze, 2018). Sustainability involves balancing the coexistence of humans and the environment for the societal well-being of current and future generations (Borowy, 2013). A common conceptualization of sustainability is comprised of three pillars, including environmental, social, and economic sustainability (Purvis, Mao, & Robinson, 2019). From an ecological standpoint, the need to live sustainably arose from the overuse and exploitation of natural resources (Robinson, 2012). This calls attention to how these resources and the damages caused by overuse are distributed, affecting the economic and societal components of sustainability (Clayton, Kals, & Feygina, 2016), as well as the availability and access to education to help solve sustainability and social justice issues (McKenzie, 2004). Environmental education (EE) is one tool that can help address these challenges. EE has been shown to have powerful impacts on youth, increasing knowledge and awareness of environmental issues, changing or enhancing attitudes as they relate to the environment, promoting environmentally responsible behaviors, and building critical thinking, leadership, and collaboration skills (Ardoin, Bowers, Roth, & Holthuis, 2018; Dettmann-Easler & Pease, 1996, 1999; Stern, Powell, & Hill, 2014; Thomas, Teel, Bruyere, & Laurence, 2019). There is a growing recognition that EE should be equitably distributed, accessible to all, and more inclusive in its content and approach. 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).

There are steps the field of EE can take to address equity issues within the field itself, including promoting equitable access to and participation in outdoor activities and EE programs. Participation in EE programs and outdoor recreation on public land has long been unequal between people of color and White people (Johnson, Bowker, English, & Worthen, 1997; Pease, 2015; Roberts, 2007; Warren, Roberts, Breunig, & Alvarez, 2014). Early explanations of these disparities suggested that people of color might have different cultural values, harbor negative perceptions of the environment, and be less interested in learning about and spending time in nature (Johnson *et al.*, 1997; Peterson, 1977). On closer inspection, many of the previous findings have been challenged. Studies have found evidence that people of color are not less interested in environmental issues, and in fact, often found themselves bearing the brunt of environmental damages (Lazri & Konisky, 2019; Macias, 2016). Further, research has shown that the foundations of the environmental field, couched in White, Eurocentric values, may have caused some of the disparities we see today (McLean, 2013; Finney, 2014; Warren *et al.*, 2014). Social and cultural norms manifest themselves by creating a narrative that the environment is White space causing people of color to feel unwelcome and discriminated against in these spaces. EE is a space that can be welcoming and relevant to all people, regardless race, ethnicity, or culture. Further investigation is needed to understand the sociocultural barriers faced by youth of color to participate in EE programs.

Evaluation is one tool that can help EE organizations improve and meet the needs of all audiences. In a rapidly changing and evolving world, the field of EE will need to embrace a culture of continual learning and improvement to ensure programming is preparing future generations with the knowledge, skills, ethics, and motivations necessary to solve the challenges we face as a global society. Evaluation is a key component of continual learning and improvement, and historically, evaluation has not been used to its fullest potential in the field of EE (Brody, Storksdieck, Stevenson, Dillon, & Wals, 2014; Carleton-Hug & Hug, 2010; Fien, Scott, & Tilbury, 2001; Tao, 2012). This may stem from a perception of evaluation in EE as an accountability measure, imposed by funding agencies to ensure their contributions were being used effectively and efficiently (Carleton-Hu & Hug, 2010; Wandersman, Imm, Chinman, & Kaftarian, 2000). This accountability mindset can cause EE practitioners to be fearful of negative evaluation results, as the results may impact their reputation and future funding opportunities (Carleton-Hug & Hug, 2010; McDuff 2002). Evaluations can be more beneficial to EE organizations than simply demonstrating successes to external stakeholders. Evaluations can be used to identify areas for improvement and provide evidence to inform programmatic adjustments (Powell, Stern, & Ardoin, 2006). With an adaptive management mindset, organizations are encouraged to be flexible and try new techniques to further reach intended goals. Here, failure or negative evaluation results simply highlights areas for modifications and improvement. The adaptive management process is also iterative, as new techniques are implemented, evaluated, and adjusted, promoting an organizational culture of continual evaluation. In contrast, within an accountability perspective, an organization may conduct a one-time summative assessment and report the results to funders or other external stakeholders. The one-time assessment will not likely provide much insight into how the organization can improve their programming for future participants, and the effort and resources expended do not benefit organizational staff or program participants.

Using evaluation to make evidence-based decisions to improve programs relies on the assumption that the evaluation data are valid. Inasmuch as the environmental field has struggled to address the disparities in access and participation among racial groups, the field of evaluation has been grappling with the validity of using traditional Eurocentric evaluations methods for diverse and multicultural groups (House, 1990; Hopson, 2003; Hood, 2004; Kirkhart, 1995; 2016; Kirkhart & Hopson, 2010; Stake, 1975). Evaluations inherently privilege some values and ways of knowing over others. Evaluators, often with input from organization staff, determine what success looks like and how it is defined, what the criteria are for measuring success, and how the data will be collected (Kirkhart, 2016). These decisions are value-laden and can be influenced by the evaluators and others who are involved in the evaluation process (House & Howe, 1999; Kirkhart, 1995). When evaluators and program staff have differing cultures, ways of knowing, and values than program participants, the evaluation may not be relevant to program participants. The field of evaluation has built up an abundance of literature to address this issue, branching into various evaluation approaches including 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). Applying these concepts to evaluation in the field of EE has the potential to uncover implicit

biases embedded within EE and its evaluation. It also can begin to address the equity issues within the field, and build a base of culturally valid evidence to re-envision and re-create the field of EE as a space for all, where all people can benefit from the work environmental educators do to increase environmental literacy and promote a sustainable and just future.

This dissertation seeks to address three current challenges within the field of EE, including:

1. increasing student attendance to EE programs and addressing the disparities in attendance between White youth and youth of color;
2. improving evaluation processes to meet the needs of EE practitioners; and
3. promoting diversity, equity, and inclusion throughout EE programming and evaluation practices.

Each chapter within this dissertation represents a stand-alone manuscript for publication. Chapter 2 investigates factors that influence student attendance to a residential EE camp with a particular emphasis on the effects of students' race/ethnicity. Chapter 3 explores EE practitioner satisfaction with evaluation processes and identifies areas and pathways for improving practitioner satisfaction and the use of evaluation in EE. Chapter 4 outlines a culturally responsive evaluation framework adapted for use withing the field of EE. This framework provides guidance on conducting evaluations that are attuned to issues of race, power, and privilege, and that are sensitive to and valid for the culture and context in which the program is located. Chapter 5 concludes the dissertation by first highlighting the cross-cutting themes of evaluation and DEI throughout these manuscripts, then providing a reflection of my experiences throughout this work, and ending with thoughts about the future of my career in academia and the field of EE.

References

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.

Balaceanu, C., Apostol, D., & Penu, D. (2012). Sustainability and social justice. *Procedia-Social and Behavioral Sciences*, *62*, 677-681.

Borowy, I. (2013). *Defining sustainable development for our common future: A history of the World Commission on Environment and Development (Brundtland Commission)*. New York, NY: Routledge.

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).

Brody, M., Storksdieck, M., Stevenson, R. B., Dillon, J., & Wals, A. E. J. (2013). Evaluation and analysis of environmental education programs, materials, and technologies and the assessment of learners and learning. In R.B. Stevenson, M. Brody, J. Dillon, & A.E.J. Wals (Ed.). *International handbook of research on environmental education* (pp. 285-288). New York, NY: Routledge.

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

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.

Clayton, S., Kals, E., & Feygina, I. (2016). Justice and environmental sustainability. In C. Sabbagh & M. Schmitt (Eds.). *Handbook of social justice theory and research* (pp. 369-386). New York, NY: Springer.

Dettmann-Easler, D., & Pease, J. L. (1996). Days of wonder. *The Science Teacher*, *63*(6), 41.

Dettmann-Easler, D., & Pease, J. L. (1999). Evaluating the effectiveness of residential environmental education programs in fostering positive attitudes toward wildlife. *The Journal of Environmental Education*, *31*(1), 33-39.

Fien, J., Scott, W., & Tilbury, D. (2001). Education and conservation: Lessons from an evaluation. *Environmental Education Research*, *7*(4), 379-395.

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.

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.

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.

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., & Howe, K. R. (1999). *Values in evaluation and social research*. Thousand Oaks, CA: Sage Publications.

Johnson, C. Y., Bowker, J. M., English, D. B., & Worthen, D. (1997). *Theoretical perspectives of ethnicity and outdoor recreation: a review and synthesis of African-American and European-American participation.* Gen. Tech. Rep. SRS-11. Asheville, NC: US Department of Agriculture, Forest Service, Southern Research Station.

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. (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.

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.

Lazri, A. M., & Konisky, D. M. (2019). Environmental attitudes across race and ethnicity. *Social Science Quarterly*, *100*(4), 1039-1055.

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.

Macias, T. (2016). Ecological assimilation: Race, ethnicity, and the inverted gap of environmental concern. *Society & Natural Resources*, *29*(1), 3-19.

McDuff, M. (2002). Needs assessment for participatory evaluation of environmental education programs. *Applied Environmental Education and Communication: An International Journal*, *1*(1), 25-36.

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

McKenzie, S. (2004). *Social sustainability: towards some definitions.* (Working Paper 27). Magill, South Australia: Hawke Research Institute.

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

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.

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

Peterson, G. L. (1977). *Recreational preferences of urban teenagers: The influence of cultural and environmental attributes.* Gen. Tech. Rep. NE-30. Upper Darby, PA: US Department of Agriculture, Forest Service, Northeast Forest Experiment Station.

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

Purvis, B., Mao, Y., & Robinson, D. (2019). Three pillars of sustainability: in search of conceptual origins. *Sustainability science*, *14*(3), 681-695.

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

Robinson, Z. (2012). *The Answers: Sustainability: How can you learn to live more sustainably, without giving up your lifestyle, and why should you care?*. Singapore: Marshall Cavendish International.

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.

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

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.

Sze, J. (Ed.). (2018). *Sustainability: Approaches to environmental justice and social power*. New Yok, NY: New York University Press.

Tao, Z. (2012). Education programs on environment. *Procedia Environmental Sciences*, *12*, 349-353.

Thomas, R. E., Teel, T., Bruyere, B., & Laurence, S. (2019). Metrics and outcomes of conservation education: a quarter century of lessons learned. *Environmental Education Research*, *25*(2), 172-192.

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

Wandersman, A., Imm, P., Chinman, M., & Kaftarian, S. (2000). Getting to outcomes: A results-based approach to accountability. *Evaluation and program planning*, *23*(3), 389-395.

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.

Chapter 2

Understanding attendance patterns of diverse students at a residential environmental education camp

Abstract

Environmental education programs (EE) can have powerful impacts on student learning, awareness, attitudes, and behaviors as they relate to the environment, as well as critical thinking, collaboration, and leadership skills. When students are given a choice to attend a residential EE program with their school, at no cost, what factors influence attendance to the program? To explore this question, pre-visit surveys were administered to sixth-grade students at seven schools invited to participate in a residential EE program at the NorthBay Adventure Center. Results indicate that logistic preparation and parental support to attend were each positively related to student attendance to the program. Students who identified as Black were less likely to attend the program compared to White students, while students of color who did not identify as Black (e.g., Latinx, mixed ethnicities) reported lower levels of parental support to attend the program compared to White students. This indicates that students' racial identities may give rise to unique barriers to program attendance. Increased communication between EE programs, classroom teachers, students, and their parents could help boost attendance, but the complex social history of discrimination in the outdoors may be at play.

Keywords: environmental education, program attendance, diversity, inclusion, students of color

Introduction

Decades of research have shown the powerful impacts environmental education (EE) can have on student learning, awareness, attitudes, and behaviors as they relate to the environment (Ardoin, Bowers, Roth, & Holthuis, 2018; Dettmann-Easler & Pease, 1996, 1999; Stern, Powell, & Hill, 2014; Thomas, Teel, Bruyere, & Laurence, 2019). Further, particularly in the case of longer experiences and residential camps, EE can improve participants' collaboration, leadership, and critical thinking skills; increase self-esteem and self-efficacy; and motivate students to work harder in school (Ardoin *et al.*, 2018; Dresner & Gill, 1994; Stern, Powell, & Ardoin, 2010; Stern *et al.*, 2014). Despite the clear benefits of EE, support for and participation in EE varies nationwide within the United States (Ruggiero, 2016). Moreover, participation in outdoor learning, EE, and residential camps can be unequal across groups with differing racial backgrounds and socioeconomic status, even when cost of attendance is mitigated (Bustamante, 2008; Chin & Phillips, 2004; Hong & Anderson, 2006). Understanding what factors influence student participation in EE experiences, with a particular emphasis on the impact students' racial identities have on participation, can help to reveal pathways forward for expanding access to EE across diverse segments of the U.S. population.

Prior literature suggests that student attendance to educational programs, field trips, and residential camps is likely influenced by personal interest, peer influence (Borden *et al.*, 2006; Grossman *et al.*, 2002), competing responsibilities (Borden *et al.*, 2006), and support and enthusiasm to attend from adults, such as parents and teachers (Chin & Phillips, 2004; Esters & Bowen, 2004; Fletcher, Elder, & Mekos, 2000; Hultsman, 1993; Jarvis & Pell, 2005; Stern,

Wright, & Powell, 2012). Additionally, students' comfort level with leaving home for an extended period of time (Thurber, 2005) and the amount of information and preparation received prior to attending (Behrendt & Franklin, 2014; Myers & Jones, 2004) can also influence a student's likelihood of attending a field trip or residential program. Many of these factors may be influenced by race, culture, and social norms. Chin and Phillips (2004) assert that social class and race tend to influence parents' ability to cultivate their children's interests and gain access to programs and activities that contribute to their children's well-being. This can significantly disadvantage children from lower socioeconomic classes and further entrench inequities in program participation (Chin & Phillips, 2004). The majority of these aforementioned studies were not explicitly examining these factors in the context of environmental education and were focused more broadly on youth programs (Borden *et al.*, 2006; Grossman *et al.*, 2002), summer activities and camps (Chin & Phillips, 2004; Thurber, 2005), and school field trips (Behrendt & Franklin, 2014; Myers & Jones, 2004). Race and culture may have a unique impact on factors that influence attendance to environmental education programs when the complex social history of discrimination in the outdoors is considered.

Literature pertaining to outdoor recreation, outdoor experiential education, and park visitation highlight several unique barriers people of color face to participate in outdoor activities including 1) access, including transportation and cost; 2) communication, including language barriers and access to information; 3) fear of discrimination; 4) a lack of knowledge and experience with the outdoors; and 5) a lack of diversity within park or program staff (Johnson, Bowker, English, & Worthen, 1997; Pease 2015; Roberts, 2007; Warren, Roberts, Breunig, & Alvarez, 2014). For example, one study found that although Latinx parents expressed interest in having their children participate in science-related programs, they were generally unaware of available programs and wanted to be able to participate alongside their children. Many parents felt that language barriers and community dynamics played a larger role in the unequal participation of Latinx youth as opposed to resources such as time, money, and transportation (Bruyere & Salazar, 2010). In a similar study, interviews with Latinx community members and parents of elementary-aged children revealed that familiarity with the local nature center was low among participants. For those who were familiar, the center was perceived as being isolated and only intended for wealthy White people (Hong & Anderson, 2006). Allison and Hibbler (2004) found evidence of several organizational barriers to participation for underserved and minority populations within parks and recreation organizations, including staff members' lack of awareness of the shifting demographics in their area, lack of competencies to work in a multicultural space, use of stereotypes, and negative attitudes about working with certain communities. These studies provide context to the historical relationships between traditionally underserved audiences, particularly people of color, and the use of outdoor areas and programs. Without assessing factors that influence attendance to EE programs in this context, EE programs will continue to reach the audiences they have traditionally served, and participation from underrepresented groups will likely not increase.

Theoretical Framework

To further understand the factors that may influence student attendance to a residential EE program, we draw on the concepts of novelty space (Orion & Hofstein, 1994) and adult support (Chin & Phillips, 2004; Esters & Bowen, 2004; Fletcher *et al*., 2000; Hultsman, 1993; Jarvis & Pell, 2005; Stern *et al.*, 2012). EE camps and field trips require students to travel to an

outdoor location, a museum, or other environmental facility that may be unfamiliar to them. Orion and Hofstein (1994) assert that each new setting contains a 'novelty space' that should be adequately addressed for maximum learning to occur. Novelty space is essentially the extent to which students are unfamiliar with the trip and its location. Reducing the novelty space prior to the visit enables students to spend more time focusing on the content of the trip and less time worrying about the unknown. In the case of residential EE camps, novelty space will likely need to be addressed due to the fact that participants spend multiple nights in an unfamiliar place that may be more rustic and natural than they are accustomed to. To reduce this novelty space to conducive levels, Orion and Hofstein (1994) suggest focusing on three types of preparation: cognitive (the content to be covered); geographical (the physical setting of the trip); and psychological (setting expectations about events and activities). This preparation can consist of specific lessons pertaining to the subject matter of the trip; videos, websites, and maps detailing its location; and in-class discussion about the types of activities that will take place during the trip. Logistical preparation, including suggestions about what students should bring on the trip and requirements to attend, is an additional factor that has been found to improve student outcomes on field trips (Kubota & Olstad, 1991; Smith-Sebasto & Cavern, 2006; Lee, Stern, & Powell, 2020). In our exploration of the literature, we have uncovered no studies examining the relative importance of these different forms of preparation for student participation in EE field trips.

In the context of EE field trips that occur during the school year, support from both parents and teachers to attend likely influence student attendance. Teachers are often under immense pressure to meet curricular goals in a timely manner and improve students' chances of succeeding on standardized tests (Ford, 2018). For students to attend off-site programming, teachers must cede valuable time in the classroom. The level of support teachers demonstrate for students to attend field trips likely varies by the teachers' degree of autonomy in decision-making, perceptions of how well the trip aligns with curricula, the length of the program, and the perceived benefits of participation (Anderson, Kisiel, & Storksdieck, 2006; Muse, Chiarelott, & Davidman, 1982; Orion, 1993; Stern *et al.*, 2012). During the school year, students spend a considerable amount of time at school with teachers, and thus perceptions of their teachers' attitudes towards and support for participation in field trips may influence students' willingness to attend the trip. For many students, an additional powerful influence on program attendance may be the support of their parents (Chin & Phillips, 2004; Esters & Bowen, 2004; Hultsman, 1993). Schools typically require parental permission for students to attend off-site programs, and often ask for chaperone volunteers (DeWitt & Storksdieck, 2008). To allow their children to leave the home for an extended period of time, particularly in the case of residential field trips, parents must have some degree of trust in the school administrators and program staff (Bustamante, 2008; Reilly, 2008; Strier & Katz, 2016). Students would likely not participate in an off-site field trip without support from their parents to attend.

The general culture and climate of students' schools might also substantially influence student attendance. Schools in the United States can vary substantially in terms of their size (i.e., the number of students who attend the school), location (i.e., urban, suburban, rural), the racial/ethnic make-up and socioeconomic status of the student body, the resources available (e.g., Title 1 status) and student to teacher ratio. These factors contribute to an overall school culture or climate, which can impact student behavior and decision-making (Gruenert & Whitaker, 2015;

Ohlson, Swanson, Adams-Manning, & Byrd, 2016). In the case of field trips, school culture, and the broader community characteristics in which the school is situated, might influence teacher and parental support for attendance in addition to setting economic, administrative, or other social constraints upon participation.

Despite a wealth of literature on factors influencing participation in outdoor activities, little empirical evidence exists to explain youth participation in EE programs and residential camp experiences. This research examines how students' school and racial/ethnic identity, pre-trip preparation, and perceived support to attend from teachers and parents influences student attendance to a 5-day residential EE program.

Study Site

This research was conducted in partnership with NorthBay, an environmental education non-profit organization located in Cecil County, Maryland on the Chesapeake Bay. For the past 15 years, NorthBay has provided 5-day residential EE programs serving over 10,000 middle school students each school year, primarily in the 6th grade. NorthBay serves a wide range of racially and socioeconomically diverse students from neighboring counties, the greater Baltimore area, and southeastern Pennsylvania.

NorthBay programming brings together EE with positive youth development. NorthBay's primary goal is to challenge students to realize they have a choice in their life paths and that every choice leads to positive or negative consequences for their futures, their communities, and the environment. Each day, the students attend two EE lessons, three activities, a night show, and participate in evening guided reflections. The EE lessons were created using the 'Investigating and Evaluating Environmental Issues and Actions' (IEEIA) model, which uses a multidisciplinary approach to guide students through investigations, considering multiple perspectives on environmental issues and appropriate actions (Hungerford, Litherland, Volk, Ramsey, & Peyton 2003) and the Better Environmental Education, Teaching, Learning, and Expertise Sharing (BEETLES) model, which uses learner-centered practices to guide student reflections (Romero, Chi, & Snow, 2015).

After the two education lessons each day, students participate in a variety of activities, including a zip-line, kayaking, ropes course, scavenger hunt, rock wall and sports games. Some of these activities are meant to push the students out of their comfort zone to build confidence and overcome fears, such as the zip-line and rock wall. Other activities, such as team sports, are offered to provide a sense of the comfort of home. After dinner, the students attend *NorthBay Live*, which is a multimedia performance that reiterates themes from the day's lessons and stresses the importance of positive decision making and individual choice. Each night the show has different elements, including videos, live musical performances, comedy, and skits. The content integrates relevant pop culture to further engage the students. After *NorthBay Live*, the students head to their cabins for the night where a counselor leads guided reflections.

NorthBay programming was developed specifically to relate to the lives of racially diverse students from urban environments. About one-third of the educational staff at NorthBay identified as a person of color at the time of the research. Typically, all schools that plan to attend NorthBay schedule a parent information night where parents are invited to attend a short

presentation about NorthBay, and NorthBay staff are available to answer questions and concerns. Some schools also receive educational programming in school before and after the residential camp experience. In these cases, NorthBay educators travel to the school and lead lessons and discussions about the environment. These efforts were designed to ease concerns about attending camp, to build trust with the students and their parents, and to improve student learning and other outcomes.

All costs to attend NorthBay are provided by external grants or are funded by the school, and all meals are provided for the days the students are at camp. Students and their families thus do not pay to attend the program. Rather, the students and their parent or guardian must fill out and return a permission form and medical form. Teachers and parents attending as chaperones are also provided with meals and lodging and do not pay for their attendance.

Methods

Sample

Survey data were collected from 1,150 sixth-grade students from seven schools within three school districts (one in Maryland and two in Pennsylvania) between November of 2019 and March of 2020. Table 2-1 shows a breakdown of the characteristics of each school. All schools in this sample received a parent information night prior to the residential experience. Schools in Districts 2 and 3 also participated in two in-class educational lessons led by NorthBay staff before the trip. This sample was chosen to obtain responses from students of varying racial identities.

Table 2-1. Characteristics of schools included in the sample.

School	District	State	School-wide racial majority	% eligible for free/ reduced lunch	Locale	# of pre-visit educational lessons	Survey timing (# of days prior to trip)
1	1	MD	White	58%	Rural- Distant	0	17
2	1	MD	White	53%	Rural- Town	0	4
3	2	PA	Hispanic	100%	Urban- Small City	2	10
4	2	PA	Hispanic	100%	Urban- Small City	2	10
5	2	PA	Hispanic	100%	Urban- Small City	2	4
6	2	PA	Hispanic	100%	Suburb- Large	2	23
7	3	PA	No majority	100%	Suburb- Midsize	2	5

Survey

An online pre-experience survey was administered to students between four and 23 days[1] before the school was scheduled to attend NorthBay. Students were also administered a post-experience survey on the last day of the field trip. Completion of an on-site post-experience survey was used to indicate that the student attended NorthBay. Other data obtained from the post-experience surveys are not relevant to this manuscript and thus are not reported.

[1] Surveys were planned to be administered seven days before the schools were scheduled to attend NorthBay, but logistical challenges with the schools prevented consistent timing. Because of this unplanned variation, we tested the influence of survey timing on each of the variables in the statistical model.

Pre-experience survey items and the constructs measured are presented in Table 2-2. We used Orion and Hofstein's (1994) novelty space framework to categorize student knowledge that reflects the types of preparation they received in addition to measuring logistical preparation, and parent and teacher support to attend, as described above.

Table 2-2. Pre-experience survey items.

Question Stem and *Response Options*	**Survey Item**	**Construct**
How much do you know about the following things related to NorthBay? *Nothing at all (1), Hardly anything (2), A little (3), A lot (4)*	• What to bring to NorthBay • The paperwork I need to turn in to go to NorthBay	Logistical preparation
	• What the cabins will be like • What the NorthBay campus looks like	Geographic preparation
	• The kinds of activities I will be participating in • What my daily schedule will be like at NorthBay • The rules at NorthBay	Psychological preparation
	• What I will be learning at NorthBay	Cognitive preparation
On a scale from 0-10, how much do you agree with the following statements? *Scale anchors: Not at all (0), Strongly agree (10)*	• My teacher wants me to go to NorthBay	Teacher support
	• My parent or guardian wants me to go to NorthBay	Parental support

Analyses

We analyzed the data using Stata 16 software. Before running any analyses, we examined if the variance in timing of pre-experience survey administration across schools may have influenced survey responses by calculating the Pearson correlation coefficient between the number of days the survey was administered before the trip and the variables measured in the pre-experience survey. When these tests yielded no statistically significant relationships, we developed a hypothesized generalized structural equation model (GSEM) to determine what factors most powerfully influence student attendance to NorthBay. Because each student is nested within a school, a hierarchical structural model would be appropriate for this analysis; however, we were unable to obtain the minimum number of schools necessary to run this analysis (see Hox, Moerbeek, & Van de Schoot, 2017). Rather, we included students' school as a binary variable in the GSEM. We hypothesized that students' school likely influences every variable in the model. Additionally, we believed that preparation to attend (calculated by taking the mean of survey items that measure each preparation type), support from teachers and parents, and students' race would influence whether they attended NorthBay. Moreover, we predicted that students' race/ethnicity would also predict the degree of parental support they felt to attend the field trip. We first conducted a one-way ANOVA to test this pathway. Based on the results, we included it in the overall hypothesized structural equation model (Figure 2-1).

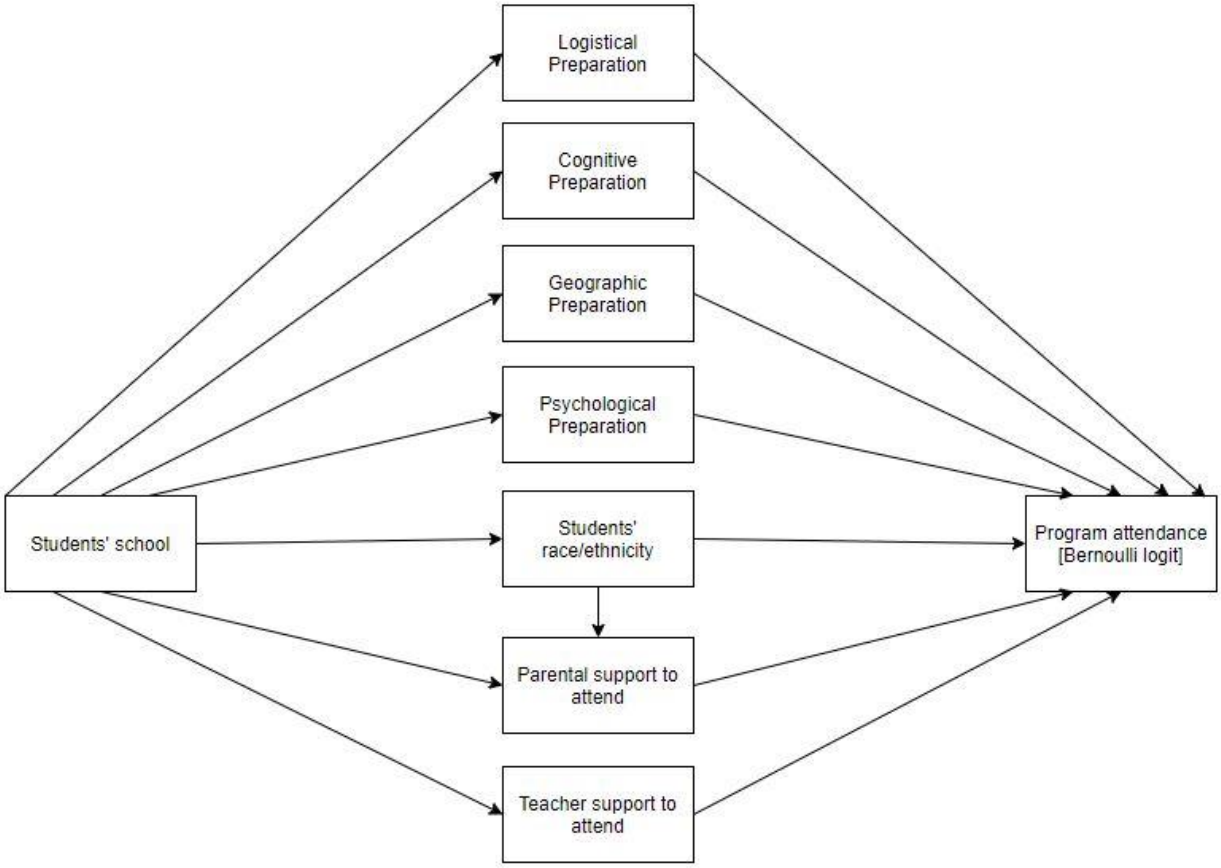

Figure 2-1: Hypothesized generalized structural equation model.

Structural equation modeling was originally developed for use with continuous endogenous (dependent) variables (Kline, 2015). The dependent (endogenous) variable in our hypothesized model is binary (program attendance). We therefore we used generalized structural equation modeling to run our analysis, which combines generalized linear modeling and structural equation modeling, allowing for non-continuous variables (Barmar *et al.*, 2018). Model estimation was conducted in Stata using a quasi-maximum likelihood estimation method and listwise deletion of missing responses (Acock, 2013). Each path or arrow in the structural model represents a regression analysis. The type of regression is determined by the structure of the dependent variable. We treated the four preparation variables and the parental and teacher support variables as continuous data; therefore, multiple linear regression was used to determine the effect of students' school on these variables. For variables regressed on program attendance, a binary variable, Bernoulli logistic regression was used. When categorical variables are included in a regression analysis, one of the categories within the variable serves as a reference group to which the other categories or levels are compared (Sperandei, 2014; Stolzfus, 2011). Without a reference group, the model suffers from collinearity and fails to converge (Midi, Sarkar, & Rana, 2010). Our structural model includes a White racial identity and School 2 as reference groups.

After running our hypothesized model, we examined the coefficients and significance of the predictors to adjust the model by removing paths that were not statistically significant. In

total, we tested four different models. Comparative model fit was determined using the Bayesian Information Criteria (BIC) (Kline, 2015; Lin, Huang & Weng, 2017) as well as theoretical meaningfulness and model parsimony (Gezie *et al*., 2018). We used the comparative fit measure, as there is no model-level goodness-of-fit tests suitable for use in GSEM (Barmar *et al*., 2018). To get a general sense of the global fit for this model, we ran a logistic regression model with program attendance as the dependent variable to obtain pseduo-R^2 values. The pseudo-R^2 value indicates how well the model is performing compared to a null model with no predictors (McFadden, 1974). This regression model included the predictors of program attendance that emerged from our final GSEM.

Results

While 72% of eligible students completed a pre-experience survey, only 57% of these students attended the NorthBay program with their school (Table 2-3). Nearly 70% of all pre-experience survey responses came from students of color. Pearson correlation coefficients indicated no statistically significant correlations between survey timing and students' reported preparation, parental support, and teacher support ($p > 0.10$).

Table 2-3. Survey responses by school.

School	Full enrollment	Survey	Survey responses	Racial/ethnic make-up of survey respondents				
				Black	Latinx	White	Another Race	Mixed Race
1	149	Pre	121 (81%)	21%	2%	60%	7%	10%
		Post	95 (64%)	20%	1%	63%	5%	11%
2	308	Pre	272 (88%)	8%	15%	55%	12%	10%
		Post	226 (73%)	7%	12%	59%	4%	14%
3	186	Pre	127 (68%)	16%	24%	18%	13%	29%
		Post	98 (53%)	14%	21%	22%	10%	32%
4	220	Pre	134 (61%)	15%	38%	5%	9%	33%
		Post	71 (32%)	13%	33%	4%	6%	44%
5	260	Pre	161 (62%)	15%	22%	10%	10%	43%
		Post	118 (45%)	12%	23%	12%	9%	43%
6	222	Pre	165 (74%)	14%	26%	19%	14%	27%
		Post	115 (52%)	12%	25%	21%	13%	29%
7	250	Pre	170 (68%)	21%	7%	26%	9%	37%
		Post	179 (72%)	22%	4%	25%	8%	42%
Total	1,595	Pre	1,150 (72%)	15%	18%	30%	11%	26%
		Post	902 (57%)	14%	15%	34%	9%	28%

Overall, students were most prepared logistically and least prepared geographically (Table 2-4). On average, students felt they had high levels of support from both teachers and parents to attend the program, both scoring above the midpoint on a 0-10 scale.

Table 2-4. Descriptive statistics of survey constructs.

Preparation (1 to 4)	**Mean**	**SD**	**n**
Logistical preparation	3.26	0.87	1,098
Cognitive preparation	2.81	1.06	1,082
Psychological preparation	2.64	0.91	1,095
Geographic preparation	2.42	0.97	1,097
Preparation (0 to 10)	**Mean**	**SD**	**n**
Teacher support	8.50	2.64	1,081
Parental support	7.54	3.29	1,034

Students from School 4 had lower levels of logistical preparation and parental support to attend, on average, while students from School 7 had higher levels of cognitive preparation, psychological preparation, and teacher support, on average (Table 2-5). Students from School 1 also had higher levels of reported teacher support.

Table 2-5. ANOVA results of independent variables by school.

	Mean variable score						
Variable	**School 1**	**School 2**	**School 3**	**School 4**	**School 5**	**School 6**	**School 7**
Logistical preparation	3.21	$3.40^{\mathbf{3,4}}$	$3.05^{\mathbf{2,7}}$	$2.84^{\mathbf{2,5,6,7}}$	$3.35^{\mathbf{4}}$	$3.25^{\mathbf{4}}$	$3.47^{\mathbf{3,4}}$
Cognitive preparation	2.85	$2.76^{\mathbf{7}}$	$2.62^{\mathbf{7}}$	2.83	$2.75^{\mathbf{7}}$	$2.64^{\mathbf{7}}$	$3.18^{\mathbf{2,3,5,6}}$
Psychological preparation	$2.65^{\mathbf{7}}$	$2.59^{\mathbf{7}}$	$2.38^{\mathbf{7}}$	$2.65^{\mathbf{7}}$	$2.49^{\mathbf{7}}$	$2.63^{\mathbf{7}}$	$3.04^{\mathbf{1,2,3,4,5,6}}$
Geographic preparation	2.41	$2.30^{\mathbf{7}}$	2.40	2.45	2.34	2.47	$2.62^{\mathbf{2}}$
Teacher support	$8.98^{2,4,6}$	$8.19^{1,\mathbf{7}}$	8.63	$7.90^{1,\mathbf{7}}$	8.83	$8.01^{1,\mathbf{7}}$	$9.15^{\mathbf{2,4,6}}$
Parental support	$8.57^{3,\mathbf{4}}$	$7.61^{\mathbf{4}}$	7.23^{1}	$5.76^{\mathbf{1,2,5,6,7}}$	$7.81^{\mathbf{4}}$	$7.85^{\mathbf{4}}$	$7.70^{\mathbf{4}}$

Superscript indicates significance at p<0.05. Bold superscript indicates significance at p<0.01.

To inform our hypothesized structural model, we conducted a one-way ANOVA between students' race and parental support (Table 2-6). The results indicated that White students reported the highest levels of parental support, while Latinx students and students of mixed race reported the lowest levels of parental support to attend the program. To capture this relationship in our structural model, we hypothesized potential mediation effects between students' race, parental support, and program attendance.

Table 2-6. ANOVA results of parental support by student's self-identified race.

	Black (B)	**Latinx (L)**	**White (W)**	**Another race (A)**	**Mixed Race (M)**	**ANOVA**
	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	F (df) p
Parental support	7.68 (3.23)	6.59 (3.62)	8.30 (2.90)	7.32 (3.48)	7.41 (3.22)	8.27 (4, 976) p<0.01
Post-hoc	B > L*	L < B*, W***	W > L***, M**		M < W**	

* p<0.05, **p<0.01, ***p<0.001. Dunnett's T3 post-hoc test used to determine significant differences. Eta-squared = 0.03

We began model testing by first examining the path coefficients and their statistical significance in our hypothesized model while also calculating the BIC value for model comparison (Table 2-7). The hypothesized model indicated no significant relationship between cognitive, psychological, and geographic preparation and program attendance. We then estimated a second model in which we removed these novelty space factors from the model. This second model had a lower (better) BIC value compared to our hypothesized model. We further adjusted the model by removing teacher support, as this variable was not a significant predictor of program attendance. Our third model had a lower BIC value compared to our second model. Finally, we removed the direct effect of students' school on program attendance, which created a better, more parsimonious model according the BIC. This led us to our fourth and final model (Figure 2-2) with all paths remaining statistically significant.

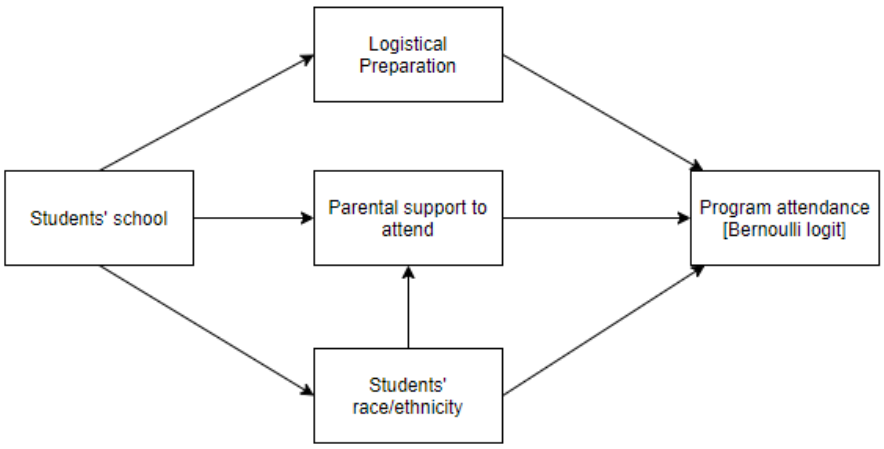

Figure 2-2. Final generalized structural equation model.

We report the exponentiated coefficients (exp(b)) of model estimates in Table 2-7. For variables regressed on program attendance, the exp(b) represents the odds ratio. For variables regressed on preparation, parental support, and teacher support, which were treated as continuous variables, the exp(b) represents the standardized beta coefficient. All of these coefficients can be interpreted in the same way, where an exp(b) of less than 1 indicates a negative relationship and an exp(b) of greater than 1 indicates a positive relationship. The results of our final model (Model 4) suggest that the level of logistical preparation students received, students' perceived level of parental support, and students' race/ethnicity significantly varied by school. Students were more likely to attend NorthBay when they felt they had higher degrees of logistical preparation and support from parents or guardians to attend. Black students were significantly less likely to attend the program compared to White students. Additionally, Latinx students, students of mixed race or ethnicities, and students who identified as a race/ethnicity other than White, Black, or Latinx reported lower levels of parental support compared to White students.

Table 2-7. Generalized structural equation model results.

		Model 1		Model 2		Model 3		Model 4	
		(BIC= 26,521)		(BIC= 17,283)		(BIC= 12,095)		(BIC= 12,061)	
Endogenous Variable	**Exogenous Variable**	**exp(b)**	**p-value**	**exp(b)**	**p-value**	**exp(b)**	**p-value**	**exp(b)**	**p-value**
Program attendance	Logistical preparation	2.01	<0.01	1.87	<0.01	1.91	<0.01	1.93	<0.01
	Cognitive preparation	1.02							
	Geographic preparation	0.94							
	Psychological preparation	0.88							
	Parental support	1.37	<0.01	1.37	<0.01	1.38	<0.01	1.37	<0.01
	Teacher support	1.04		1.04					
	Latinx	0.88		0.85		0.83		0.72	
	Black	0.64		0.64		0.61		0.58	0.03
	Other race	0.86		0.80		0.78		0.71	
	Mixed race	1.40		1.36		1.33		1.22	
	School 1	0.98		0.94		1.00			
	School 3	0.93		0.92		0.99			
	School 4	0.77		0.76		0.80			
	School 5	0.80		0.82		0.89			
	School 6	0.50	0.02	0.50	0.01	0.52	0.02		
	School 7	0.96		0.92		0.97			
Logistic preparation	School 1	0.83	0.04	0.83	0.04	0.83	0.04	0.83	0.04
	School 3	0.71	<0.01	0.71	<0.01	0.71	<0.01	0.71	<0.01
	School 4	0.57	<0.01	0.57	<0.01	0.57	<0.01	0.57	<0.01
	School 5	0.95		0.95		0.95		0.95	
	School 6	0.96		0.96		0.96		0.96	
	School 7	1.07		1.07		1.07		1.07	
Cognitive preparation	School 1	1.10							
	School 3	0.87							
	School 4	1.08							

	School 5	0.99							
	School 6	0.89							
	School 7	1.53	<0.01						
Geographic preparation	School 1	1.11							
	School 3	1.11							
	School 4	1.17							
	School 5	1.05							
	School 6	1.19							
	School 7	1.38	<0.01						
Psychological preparation	School 1	1.07							
	School 3	0.81	0.03						
	School 4	1.06							
	School 5	0.91							
	School 6	1.04							
	School 7	1.57	<0.01						
Students' race/ethnicity: Latinx	School 1	0.11	0.01	0.11	0.01	0.11	0.01	0.11	0.01
	School 3	5.10	<0.01	5.10	<0.01	5.10	<0.01	5.10	<0.01
	School 4	28.3	<0.01	28.3	<0.01	28.3	<0.01	28.3	<0.01
	School 5	8.03	<0.01	8.03	<0.01	8.03	<0.01	8.03	<0.01
	School 6	5.08	<0.01	5.08	<0.01	5.08	<0.01	5.08	<0.01
	School 7	0.97		0.97		0.97		0.97	
Students' race/ethnicity: Black	School 1	2.63	<0.01	2.63	<0.01	2.63	<0.01	2.63	<0.01
	School 3	6.44	<0.01	6.44	<0.01	6.44	<0.01	6.44	<0.01
	School 4	20.3	<0.01	20.3	<0.01	20.3	<0.01	20.3	<0.01
	School 5	10.3	<0.01	10.3	<0.01	10.3	<0.01	10.3	<0.01
	School 6	5.18	<0.01	5.18	<0.01	5.18	<0.01	5.18	<0.01
	School 7	5.66	<0.01	5.66	<0.01	5.66	<0.01	5.66	<0.01
Students' race/ethnicity:	School 1	0.53		0.53		0.53		0.53	

Other race	School 3	3.40	<0.01	3.40	<0.01	3.40	<0.01	3.40	<0.01
	School 4	7.56	<0.01	7.56	<0.01	7.56	<0.01	7.56	<0.01
	School 5	4.25	<0.01	4.25	<0.01	4.25	<0.01	4.25	<0.01
	School 6	3.28	<0.01	3.28	<0.01	3.28	<0.01	3.28	<0.01
	School 7	1.48		1.48		1.48		1.48	
Students' race/ethnicity: Mixed race	School 1	0.92		0.92		0.92		0.92	
	School 3	8.37	<0.01	8.37	<0.01	8.37	<0.01	8.37	<0.01
	School 4	33.1	<0.01	33.1	<0.01	33.1	<0.01	33.1	<0.01
	School 5	22.2	<0.01	22.2	<0.01	22.2	<0.01	22.2	<0.01
	School 6	7.21	<0.01	7.21	<0.01	7.21	<0.01	7.21	<0.01
	School 7	7.42	<0.01	7.42	<0.01	7.42	<0.01	7.42	<0.01
Parental support	School 1	2.29	0.03	2.29	0.03	2.29	0.03	2.29	0.03
	School 3	1.06		1.06		1.06		1.06	
	School 4	0.32	0.01	0.32	0.01	0.32	0.01	0.32	0.01
	School 5	1.83		1.83		1.83		1.83	
	School 6	1.90		1.90		1.90		1.90	
	School 7	1.26		1.26		1.26		1.26	
	Latinx	0.23	<0.01	0.23	<0.01	0.23	<0.01	0.23	<0.01
	Black	0.56		0.56		0.56		0.56	
	Other race	0.41	0.02	0.41	0.02	0.41	0.02	0.41	0.02
	Mixed race	0.45	0.01	0.45	0.01	0.45	0.01	0.45	0.01
Teacher support	School 1	2.22	0.01	2.22	0.01				
	School 3	1.55		1.55					
	School 4	0.75		0.75					
	School 5	1.90	0.02	1.90	0.02				
	School 6	0.84		0.84					
	School 7	2.62	<0.01	2.62	<0.01				

Only significant p-values at $p<0.05$ are reported

Because no global model fit metric exists for use in GSEM (Barmar *et al.*, 2018), we ran a logistic regression on program attendance to report McFadden's pseudo-R^2 value. The model included logistical preparation, parental support to attend, students' race/ethnicity, and students' race/ethnicity multiplied by parental support (to capture the effect of students' race/ethnicity on parental support) as independent variables. This model had a pseudo-R^2 value of 0.25, which means this model performs about 25% better than a null model (McFadden, 1974). McFadden (1974) asserts that a pseudo R^2 value of between 0.2 and 0.4 indicates good model fit. While not a perfect measure of global fit for our final structural model, this gives us an additional indication that the model performs reasonably well.

Discussion

Our results indicate that students were more likely to attend NorthBay if they had higher levels of logistical preparation and parental support to attend. Additionally, students' racial identities influenced their likelihood of attending the program both directly, and indirectly, through parental support. Black students were less likely to attend the program compared to White students, and students of color who did not identify as Black reported lower levels of parental support. This indicates that students' racial identities may give rise to unique barriers to program attendance, and that these barriers are heterogenous across racial groups. The school students attended significantly influenced these predictors of program attendance. Each school had a unique racial/ethnic make-up and varying levels of logistical and parental support. School culture and the communities in which students live should also be considered when working to increase EE program attendance. Below, we suggest pathways forward for promoting participation in EE experiences like NorthBay, particularly for students of color.

School culture

Each school has unique norms, values, structures, and leadership that shape the culture of the school (Duke, 1990; Galván, Spatzier, & Juvonen, 2011). School culture has been shown to impact students' academic achievement (MacNeil, Prater, & Busch, 2009), empathy and social behavior (Barr & Higgins-D'Alessandro, 2007), physical activity (Rickwood, 2013), and truancy (Ohlson *et al.*, 2016). Students' school can also serve as a predictor of the students' race/ethnicity and socioeconomic status, particularly in the case of public schools where geographic zoning dictates the school at which students attend (Richards, 2014). These varying school characteristics likely influence the success of strategies used to promote program attendance. Below we discuss the implications of our findings and provide suggestions for increasing student attendance to EE programs; however, these suggestions may need to be tailored to meet the needs of each school. We encourage EE organizations to build relationships and work collaboratively with school administrators and classroom teachers to further understand how each school's unique culture might influence student program attendance.

Preparation

The novelty space theory asserts that preparing students geographically, psychologically, and cognitively can reduce the overwhelming feelings that often occur when visiting a new place, increasing students' abilities to focus and learn (Orion & Hofstein, 1994). In our study, these forms of preparation were not significantly related to student attendance to the program. Rather, we found that logistical preparation was positively related to student attendance. Logistical preparation could include a detailed packing list of what to bring, suggestions for

where to purchase or borrow needed items, clear instructions for completing and returning required paperwork (e.g., permission forms, medical forms), and the provision of transportation details (when and where the students are expected to be picked up and dropped off, or if they need to provide their own transportation). Logistical preparation differs from novelty space factors in that logistical preparation consists of what students need to know to be able to attend the trip, whereas novelty space factors focus more on preparing students for what to expect once they arrive. This could indicate that logistical preparation is a precursor to other types of preparation. That is, without proper logistical preparation, other forms of preparation may be less relevant to the students and thus do not necessarily increase attendance. While the literature supports the use of these various types of preparation for improving the success of field trips (Gennaro, 1981; Lee *et al.*, 2020; Orion & Hofstein, 1994; Stern, Powell, & Ardoin, 2008), logistical preparation may be foundational in giving students the information needed to attend the program.

Parental support to attend

Perceived support from parents to attend the program also positively influenced student attendance to NorthBay. EE organizations that work to build stronger relationships with students' parents may have more success at promoting student attendance. Students of color, particularly Latinx and students of mixed ethnicities, tended to report lower levels of parental support compared to their White counterparts. This may stem from a lack of familiarity with and trust of camp and field trip experiences. Garst and colleagues (2016) found that parents tended to be more anxious about sending their children to camp if they did not have previous camp experiences themselves. In the United States, residential camps have a long history of serving predominantly White, middle to upper class audiences (Paris, 2008; Van Slyck, 2006), which likely contributes to the apprehension parents of color have about their children attending camp experiences. In our study, Latinx students reported the lowest levels of parental support. While we did not specifically ask about students' immigration status, Bustamante (2008) found that immigrant parents tended to prefer day-trips to residential camps, citing a fear of the unknown, safety, and a discomfort of having their children away at night as contributing to their apprehensions. This could indicate that immigrant families have unique concerns about sending their children to a residential camp that have not been adequately addressed.

Previous research has shown that parent interest in participating in educational programs alongside their children is a common norm in many Latinx communities (Bruyere & Salazar, 2010; Bustamante, 2008). Although parents may be welcome to attend residential programs as chaperones, many parents are unable to do so during the work week. To help families overcome this barrier, providing additional support in the form of stipends, flexible policies on chaperone attendance, and providing shorter or alternative experiences that better fit within the confines of parent schedules may help to improve program attendance in certain communities. Providing this support may help shift parents' attitudes toward the program, as the costs of having their children attend no longer outweigh the benefits.

Students' racial/ethnic identity

Our results indicated that students' racial/ethnic identity influenced program attendance, both indirectly, as students of color who did not identify as Black reported lower levels of parental support, and directly, as Black students were less likely to attend NorthBay compared to

White students. This may indicate that sociocultural barriers not measured in this study may be causing students of color, in particular Black students, to be apprehensive about attending a program like NorthBay. These barriers could include the predominantly White values that have historically shaped the foundation of the environmental field, and fears of discrimination in the outdoors (Finney, 2006, 2014; Roberts, 2007; Roberts & Chitewere, 2011). The idea that experiencing and exploring the outdoors is a primarily White activity is perpetuated by a lack of racial representation in media and advertisement of outdoor areas and activities (Martin, 2004). Although NorthBay designed their programming to resonate with urban youth from diverse backgrounds, hosts parent information sessions, provides translators for teachers, parents, and students when needed, and has a diverse staff that often reflect the communities they serve, this was not enough to reduce disparities in attendance between White and non-White students. These practices are likely necessary but insufficient to encourage participation from students of color.

In her book, *The Art of* Relevance, Nina Simon (2016) describes ways in which organizations can not only prevent exclusion, but also promote inclusion and increase their program's relevance to culturally diverse audiences. She asserts it is not enough to open the existing door wider. Rather, she advocates for creating new doors that are relevant and enticing to underserved audiences. This means that instead of modifying existing programming and communication efforts to reach these audiences, a more successful strategy may be working with underserved communities to reimagine and co-create programming from the ground up to reflect the lived experiences of those communities. Relevance is a constantly moving target and heterogeneous within and between different racial groups. As our results showed, each school group is unique, with different racial make-ups, culture, teacher support and parent involvement. When working in a multicultural space, using a single curriculum or program structure for all school groups may not be a recipe for success. Working with each community and school to build a relevant experience that will leave them feeling welcomed and included could help increase program participation of students of color and move the field toward a more equitable future.

Limitations

Two of the seven schools included in our sample did not receive a pre-trip educational lesson; however, we were not able to isolate the effect these lessons might have had on student attendance. The absence of the pre-trip lesson is confounded by students' race, as the two schools that did not received these lessons were the only two majority White schools in the sample. While we could not isolate this effect, we did examine the potential influence of receiving this lesson on attendance to NorthBay. Results showed that students who did not receive a pre-visit educational lesson (the two majority White schools) were more likely to attend NorthBay ($p=0.01$; $V=0.07$). The Cramer's V value indicates a very small effect size[2], indicating this result may not be meaningful. We cannot confidently report the possible impact that the pre-trip educational lesson had on student attendance to NorthBay, and future research may be needed to determine the efficacy of pre-trip lessons on increasing student attendance, particularly among underserved audiences.

[2] A Cramer's V value of 0.1-0.3 is considered a small effect; 0.3-0.5 a medium effect; and >0.5 a large effect.

The time at which the pre-experience survey was implemented may have affected these results. While we had planned on surveying all students 7 days prior to their trip to NorthBay, logistical challenges and exam and holiday schedules prevented consistent timing of implementation. Our results indicated that logistical preparation was a key factor influencing student attendance to the program. Schools that administered the student survey closest to the day of the field trip, including Schools 2 and 5 (four days prior) and School 7 (five days prior) had the highest levels of logistical preparation. It is possible that students who took the survey ten days or more in advance of the trip received additional preparation prior to attending that was not captured in this study.

Collecting data from only one EE program limits the generalizability of this study; however, there are many residential EE camps around the country that are similar in structure, content, or pedagogy used at NorthBay. This might include other residential camps, camps that focus on both EE and positive youth development, and camps that incorporate the IEEIA model (Hungerford *et al.,* 2003), the BEETLES model (Romero *et al.,* 2015), or other similar approaches into their program design. We provide suggestions for these and other program providers to consider in light of their own contexts. We urge researchers to conduct similar investigations in other settings to extend, contest, and expand the field's knowledge in this area.

Conclusion

This study indicates that frequent, open, and trustworthy communication between program providers, students, and their parents may help to boost student attendance, particularly students of color. Students who felt their parents supported their attendance and felt better logistically prepared were most likely to attend the NorthBay program. Particular attention should be paid to how students' racial identities and cultures may give rise to unique barriers to program attendance including parental support. Ultimately, to increase the relevance, cultural accessibility, and student attendance to EE programs for all youth, programs may need to be flexible in content and structure, and be co-created with each community served.

References

Acock, A. C. (2013). *Discovering structural equation modeling using Stata* (Revised ed.). College Station, TX: Stata Press Books.

Allison, M. T., & Hibbler, D. K. (2004). Organizational barriers to inclusion: Perspectives from the recreation professional. *Leisure Sciences*, *26*(3), 261-280.

Anderson, D., Kisiel, J., & Storksdieck, M. (2006). Understanding teachers' perspectives on field trips: Discovering common ground in three countries. *Curator: The Museum Journal*, *49*(3), 365-386.

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.

Barmar, S., Alimohammadian, M., Sadjadi, S. A., Poustchi, H., Hosseini, S. M., & Yasseri, M. (2018). Generalized Structural Equation Modeling (GSEM) and its Application in Health Researches. *Journal of School of Public Health and Institute of Public Health Research*, *16*(1), 51-62.

Barr, J. J., & Higgins-D'Alessandro, A. (2007). Adolescent empathy and prosocial behavior in the multidimensional context of school culture. *The Journal of Genetic Psychology*, *168*(3), 231-250.

Behrendt, M., & Franklin, T. (2014). A review of research on school field trips and their value in education. *International Journal of Environmental and Science Education*, *9*(3), 235-245.

Borden, L. M., Perkins, D. F., Villarruel, F. A., Carleton-Hug, A., Stone, M. R., & Keith, J. G. (2006). Challenges and opportunities to Latino youth development: Increasing meaningful participation in youth development programs. *Hispanic Journal of Behavioral Sciences*, *28*(2), 187-208.

Bruyere, B. L., & Salazar, G. (2010). Engaging Latino audiences in out-of-school programs about science. *Journal of Extension*, *48*(3), 3-8.

Bustamante, K. (2008). *Whose nature?: Exploring the link between wilderness, belonging and residential summer camp use among Canadian immigrants* (Unpublished master's thesis). Ryerson University, Toronto, ON. Retrieved from http://digitalcommons.ryerson.ca

Chin, T., & Phillips, M. (2004). Social reproduction and child-rearing practices: Social class, children's agency, and the summer activity gap. *Sociology of education*, *77*(3), 185-210.

Dettmann-Easler, D., & Pease, J. L. (1996). Days of wonder. *The Science Teacher*, *63*(6), 41.

Dettmann-Easler, D., & Pease, J. L. (1999). Evaluating the effectiveness of residential environmental education programs in fostering positive attitudes toward wildlife. *The Journal of Environmental Education*, *31*(1), 33-39.

DeWitt, J., & Storksdieck, M. (2008). A short review of school field trips: Key findings from the past and implications for the future. *Visitor studies*, *11*(2), 181-197.

Dresner, M., & Gill, M. (1994). Environmental education at summer nature camp. *The Journal of Environmental Education*, *25*(3), 35-41.

Duke, D. L. (1990). School organization, leadership, and student behavior. *Student discipline strategies: Research and practice*, 19-46.

Esters, L. T., & Bowen, B. E. (2004). Factors Influencing Enrollment in an Urban Agricultural Education Program. *Journal of Career and Technical Education*, *21*(1), 25-37.

Finney, C. M. (2006). *Black faces, white spaces: African Americans and the great outdoors*. Worchester, MA: Clark University.

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.

Fletcher, A. C., Elder, G. H. J., & Mekos, D. (2000). Parental influences on adolescent involvement in community activities. *Journal of Research on Adolescence*, *10*(1), 29-48.

Ford, A. J. (2018). *The impact of excessive high-stakes standardized tests on students and teachers* (Doctoral dissertation.). Retrieved from ProQuest Dissertations and Theses database. (No. 10812521).

Galván, A., Spatzier, A., & Juvonen, J. (2011). Perceived norms and social values to capture school culture in elementary and middle school. *Journal of Applied Developmental Psychology*, *32*(6), 346-353.

Garst, B. A., Gagnon, R. J., & Bennett, T. (2016). Parent anxiety causes and consequences: Perspectives from camp program providers. *The Cyber Journal of Applied Leisure and Recreation Research*, *18*(1).

Gennaro, E. D. (1981). The Effectiveness of Using Pre-visit Instructional Materials on Learning for a Museum Field Trip Experience. *Journal of Research in Science Teaching*, *18*(3), 275-79.

Gezie, L. D., Yalew, A. W., Gete, Y. K., Azale, T., Brand, T., & Zeeb, H. (2018). Socio-economic, trafficking exposures and mental health symptoms of human trafficking returnees in Ethiopia: using a generalized structural equation modelling. *International Journal of Mental Health Systems*, *12*(1), 1-13.

Grossman, J. B., Price, M. L., Fellerath, V., Jucovy, L. Z., Kotloff, L. J., Raley, R., & Walker, K. E. (2002). *Multiple choices after school: Findings from the extended-service schools initiative.* Philadelphia, PA: Public/Private Ventures.

Gruenert, S., & Whitaker, T. (2015). *School culture rewired: How to define, assess, and transform it*. Alexandria, VA: ASCD.

Hong, A., & Anderson, D. H. (2006). Barriers to participation for Latino people at Dodge Nature Center. *The Journal of Environmental Education*, *37*(4), 33-44.

Hox, J. J., Moerbeek, M., & Van de Schoot, R. (2017). *Multilevel analysis: Techniques and applications*. New York, NY: Routledge Taylor & Francis Group.

Hultsman, W. Z. (1993). The influence of others as a barrier to recreation participation among early adolescents. *Journal of Leisure Research*, *25*(2), 150-164.

Hungerford, H. R., Volk, T. L., Ramsey, J. M., Litherland, R. A., & Peyton, R. B. (2003). *Investigating and evaluating environmental issues and actions: Designed as a critical thinking, problem solving approach to local, regional, and international issues.* Champaign, Illinois: Stipes Publishing LLC.

Jarvis, T., & Pell, A. (2005). Factors influencing elementary school children's attitudes toward science before, during, and after a visit to the UK National Space Centre. *Journal of Research in Science Teaching*, *42*(1), 53-83.

Johnson, C. Y., Bowker, J. M., English, D. B., & Worthen, D. (1997). *Theoretical perspectives of ethnicity and outdoor recreation: a review and synthesis of African-American and European-American participation.* Gen. Tech. Rep. SRS-11. Asheville, NC: US Department of Agriculture, Forest Service, Southern Research Station.

Kline, R. B. (2015). *Principles and practice of structural equation modeling* (4th ed.). New York, NY: Guilford Publications.

Kubota, C. A., & Olstad, R. G. (1991). Effects of novelty-reducing preparation on exploratory behavior and cognitive learning in a science museum setting. *Journal of Research in Science Teaching*, *28*(3), 225-234.

Lee, H., Stern, M. J., & Powell, R. B. (2020). Do pre-visit preparation and post-visit activities improve student outcomes on field trips?. *Environmental Education Research*, *26*(7), 989-1007.

Lin, L. C., Huang, P. H., & Weng, L. J. (2017). Selecting path models in SEM: A comparison of model selection criteria. *Structural Equation Modeling: A Multidisciplinary Journal*, *24*(6), 855-869.

MacNeil, A. J., Prater, D. L., & Busch, S. (2009). The effects of school culture and climate on student achievement. *International Journal of leadership in Education*, *12*(1), 73-84.

Martin, D. C. (2004). Apartheid in the great outdoors: American advertising and the reproduction of a racialized outdoor leisure identity. *Journal of Leisure Research*, *36*(4), 513-535.

McFadden, D. (1974). Conditional Logit Analysis of Qualitative Choice Behaviour. In P, Zarembka (Ed.). *Frontiers in Econometrics*. (pp. 105-142). New York: Academic Press.

Midi, H., Sarkar, S. K., & Rana, S. (2010). Collinearity diagnostics of binary logistic regression model. *Journal of Interdisciplinary Mathematics*, *13*(3), 253-267.

Muse, C., Chiarelott, L., & Davidman, L. (1982). Teachers' utilization of field trips: Prospects and problems. *The Clearing House*, *56*(3), 122-126.

Myers, B., & Jones, L. (2004). *Effective use of field trips in educational programming: A three stage approach*. Gainesville, FL: University of Florida Cooperative Extension Service, Institute of Food and Agricultural Sciences, EDIS.

Ohlson, M., Swanson, A., Adams-Manning, A., & Byrd, A. (2016). A Culture of Success-Examining School Culture and Student Outcomes via a Performance Framework. *Journal of Education and Learning*, *5*(1), 114-127.

Orion, N. (1993). A model for the development and implementation of field trips as an integral part of the science curriculum. *School Science and Mathematics*, *93*(6), 325-31.

Orion, N., & Hofstein, A. (1994). Factors that influence learning during a scientific field trip in a natural environment. *Journal of research in science teaching*, *31*(10), 1097-1119.

Paris, L. (2008). *Children's nature: The rise of the American summer camp*. New York: New York University Press.

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

Reilly, E. (2008). Parental involvement through better communication. *Middle School Journal*, *39*(3), 40-47.

Richards, M. P. (2014). The gerrymandering of school attendance zones and the segregation of public schools: A geospatial analysis. *American Educational Research Journal*, *51*(6), 1119-1157.

Rickwood, G. (2013). School culture and physical activity: a systematic review. *Canadian Journal of Educational Administration and Policy*, (143).

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.

Roberts, N. S., & Chitewere, T. (2011). Speaking of justice: Exploring ethnic minority perspectives of the Golden Gate National Recreation Area. *Environmental Practice*, *13*(4), 354-369.

Romero, V., Chi, B., & Snow, J. (2015). *BEETLES: Better Environmental Education, Teaching, Learning, and Expertise Sharing Final Evaluation Report.* Berkeley, CA: The Lawrence Hall of Science, University of California, Berkeley.

Ruggiero, K. (2016). *A Criteria-Based Evaluation of Environmental Literacy Plans in the United States.* Knoxville, TN: University of Tennessee.

Simon, N. (2016). *The Art of Relevance*. Santa Cruz, CA: Museum 2.0.

Smith-Sebasto, N. J., & Cavern, L. (2006). Effects of pre-and post-trip activities associated with a residential environmental education experience on students' attitudes toward the environment. *The Journal of Environmental Education*, *37*(4), 3-17.

Sperandei, S. (2014). Understanding logistic regression analysis. *Biochemia Medica*, *24*(1), 12-18.

Stern, M. J., Powell, R. B., & Ardoin, N. M. (2008). What difference does it make? Assessing outcomes from participation in a residential environmental education program. *The Journal of Environmental Education*, *39*(4), 31-43.

Stern, M. J., Powell, R. B., & Ardoin, N. M. (2010). Evaluating a constructivist and culturally responsive approach to environmental education for diverse audiences. *The Journal of Environmental Education*, *42*(2), 109-122.

Stern, M. J., Wright, M. E., & Powell, R. B. (2012). Motivating participation in national park service curriculum-based education programs. *Visitor Studies*, *15*(1), 28-47.

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.

Stoltzfus, J. C. (2011). Logistic regression: a brief primer. *Academic Emergency Medicine*, *18*(10), 1099-1104.

Strier, M., & Katz, H. (2016). Trust and parents' involvement in schools of choice. *Educational Management Administration & Leadership*, *44*(3), 363-379.

Thomas, R. E., Teel, T., Bruyere, B., & Laurence, S. (2019). Metrics and outcomes of conservation education: a quarter century of lessons learned. *Environmental Education Research*, *25*(2), 172-192.

Thurber, C. A. (2005). Multimodal homesickness prevention in boys spending 2 weeks at a residential summer camp. *Journal of Consulting and Clinical Psychology*, *73*(3), 555.

Van Slyck, A. A. (2006). *A manufactured wilderness: Summer camps and the shaping of American youth, 1890-1960*. Minneapolis, MN: University of Minnesota Press.

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.

Identifying areas and approaches for improving evaluation processes and practitioner satisfaction in environmental education

Abstract

Historically, evaluation has not been used to its fullest potential in environmental education (EE), possible due to the limited resources available for evaluation and practitioners harboring negative attitudes or perceptions of evaluation. Pressures from external stakeholders can cause organizations to expend their resources on reporting requirements at the expense of conducting evaluation activities that promote programmatic improvement and mission achievement. Understanding practitioners' satisfaction with their evaluation processes and the drivers of this satisfaction may reveal strategies for improving evaluation processes and practitioner satisfaction, potentially promoting more widespread use of evaluation to inform programmatic improvement. To understand the current state of practitioner satisfaction with evaluation processes, determine influential drivers of this satisfaction, and identify areas and approaches for the improvement of evaluation processes, we partnered with the North American Association for Environmental Education and the Association of Nature Center Administrators to administer an online survey to environmental education practitioners via email and social media. Our results indicate that satisfaction with current evaluation processes remains somewhat low. Focusing on achieving adaptive management goals, understanding participant satisfaction with program experiences, and using formal evaluation methods may help enhance practitioner satisfaction with evaluation processes. Additionally, our results highlight two areas that could use significant improvement: understanding and executing the adaptive management process and using evaluation to further diversity, equity, and inclusion efforts in the field.

Introduction

Evaluation can be used in environmental education (EE) to determine the extent to which EE programs achieve intended participant outcomes. These outcomes could include increasing knowledge and awareness about the environment, enhancing or changing environmental attitudes and behaviors, supporting social-emotional learning, improving critical thinking, collaboration, and leadership skills, and increasing motivations to do well in school (Ardoin, Bowers, Roth, & Holthuis, 2018; Stern, Powell, & Hill, 2014). Without evaluation, environmental educators may make the assumption that if students are enjoying the program and engaged in the activities, they are achieving one or more of the outcomes listed above (Frensley, Stern, & Powell, 2020; Ham, 2013). Sam Ham (2013) refers to this notion as 'poof theory,' meaning if participants are seemingly enjoying themselves, programmatic goals are likely being achieved; however, there is evidence to suggest these assumptions may be faulty. A study by Frensley and colleagues (2020) found that student self-reported engagement was a much better predictor of student outcomes achievement compared to observations made of student engagement and enthusiasm. This indicates that even participant outcomes like enjoyment, satisfaction, and engagement are best ascertained through systematic evaluations of the program rather than casual observation. Further, evaluation can also be used as an adaptive management tool, applying the results to make evidence-based decisions to improve programming (Powell, Stern, & Ardoin, 2006).

There are numerous approaches to evaluation. We focus here on some basic distinctions, between informal and formal evaluations, and between internal and externally-supported evaluations. Informal evaluations include discussions and reflections with staff members about

perceptions of program success and impact, informal observations of program implementation, and reviewing general feedback from participants and other stakeholders. What sets informal evaluation apart from formal evaluation is that informal evaluation does not include the systematic collection and analysis of data (Stufflebeam & Shinkfield, 2012). Formal evaluation implements systematic data collection and analysis procedures following best practices in research (Mertens & Wilson, 2018; Stufflebeam & Shinkfield, 2012). Formal evaluations tend to be more complex and produce data considered more scientifically valid than informal evaluations (Davidson, 2005). Because of this added complexity, some organizations partner with external evaluation professionals to help them conduct formal evaluations of their programs, or hire external evaluation professionals to conduct the evaluation themselves and report back on the findings.

Evaluations of EE programs can achieve multiple goals, ranging from satisfying accountabilities to external audiences to achieving adaptive management goals (Capwell, Butterfoss, & Francisco, 2000; Powell *et al.*, 2006; Stufflebeam & Shinkfield, 2012). Evaluations conducted to satisfy external accountabilities may focus on reporting successes to funders, organizational leadership, advisory boards, or other supporting bodies. They might also be conducted to report outcomes achievement to program participants and community audiences to increase program attendance. Evaluations for adaptive management purposes, on the other hand, prioritize organizational learning (Powell *et al.*, 2006). In its simplest form, adaptive management in the context of EE includes designing a program and its evaluation, implementing the program and evaluation, and making program adjustments based on the evaluation findings (Figure 3-1). This process is continuous, meaning what is learned from the evaluation is implemented as an adjustment to the program or the creation of a new program. These adjustments are then implemented and evaluated, leading to more learning and adjusting. Organizations may take an adaptive management approach to examine how their programs affect different audiences to better promote diverse, equitable, and inclusive participation (Cordie, 2009; Stern, Powell, Ardoin, 2010).

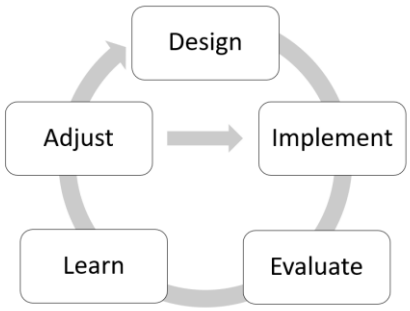

Figure 3-1. The adaptive management cycle.

The literature reveals tensions between evaluations conducted for the purpose of external accountability and those conducted as part of an adaptive management process (West, Schultz, & Bekessy, 2016). While complying with external accountabilities are important, particularly if an organization relies on external funding, placing too much focus on external accountabilities when planning for an evaluation can limit the utility of the evaluation products (Stern, 2018). EE practitioners may harbor negative attitudes and perceptions about evaluation when external accountabilities are the main, or even sole purpose for conducting an evaluation, and when the

evaluation goals align with what external stakeholders expect rather than providing organizations with actionable information to improve programs. Alternatively, prioritizing adaptive management when designing the evaluation process can help organizations clarify programmatic goals and directly link evaluation outcomes to mission achievement and programmatic improvement (Powell *et al.*, 2006). Adaptive management involves some level of uncertainty and risk (Aceves-Bueno *et al.*, 2015) and typically involves a longer process than reporting on monitoring metrics or gathering data on participant satisfaction. These uncertain longer-term outcomes may not be prioritized over pressing and immediate short-term reporting requirements (Carleton-Hug & Hug, 2010; Heimlich, 2010; Keene & Blumstein, 2010). With limited resources, the pressure to comply with a myriad of accountabilities and reporting requirements can lead to staff burnout and negative perceptions of evaluation (McDuff, 2002).

While these studies point to reasons why practitioners may have negative attitudes towards evaluation, they did not empirically assess practitioners' satisfaction with their organization's evaluation processes. Satisfaction in this context can be conceptualized as practitioners' overall affective assessment of their evaluation processes and the extent to which these processes meet practitioners' needs (Giese & Cote, 2000). Understanding practitioner satisfaction with their evaluation processes, and the drivers of this satisfaction may be critical, as practitioners that are not satisfied with their evaluations may harbor negative perceptions about evaluation, which likely hinders its use for evidence-based decision making and programmatic improvement. The multiple ways in which evaluations are conducted in the field of EE may lead to differing levels of practitioner satisfaction with their evaluation processes. This could include informal versus formal evaluations, internal versus supported evaluations, as well as the tensions to comply with external accountabilities versus using evaluation as an adaptive management tool. Understanding how the formality, level of external support, and the achievement of various evaluation goals influence practitioner satisfaction could help to identify paths forward for increasing the potential use of evaluation in EE. More specifically, our research questions include:

1. How satisfied are EE practitioners with their evaluation processes?
2. What drives EE practitioner satisfaction with their evaluation processes?
3. How might evaluation processes in EE be improved to better achieve evaluation goals and increase practitioner satisfaction with their evaluation processes?

Methods

Sample

To reach a wide range of EE practitioners, we partnered with the North American Association for Environmental Education (NAAEE) and the Association of Nature Center Administrators (ANCA) to administer an online survey via email to all members between June and July of 2020. The survey was also promoted on social media platforms. This type of non-random sample with no specific sampling frame can increase the chances of obtaining a biased sample (Dillman, Smyth, & Christian, 2014). In this case, there is likely bias toward survey respondents that are more engaged in conducting program evaluations and are more active on social media and in their respective professional associations (NAAEE and ANCA).

Survey

To adequately describe our non-random sample, we collected detailed information on respondents' roles within their organization, geographic location, organization type, organization size, program types, funding sources, and audiences served. We also collected information on the types of evaluations being conducted (i.e., formal and informal) and the personnel who conduct the evaluations (i.e., internal, externally supported, or fully external) (Table 3-1). We defined formal evaluation as "*the systematic collection and analysis of data to draw conclusions and make informed decisions about the effectiveness of programs."* Informal evaluation, on the other hand, was defined as involving "*some level of reflection and assessment, though it does not entail systematic methods for data collection and analysis. It can include reviewing and discussing programming based on the observations and thoughts of staff, participants, or others. It could also include informal (non-systematic) reviews of information collected about programs and participants*." Internal evaluations were defined as evaluations in which internal staff member(s) of the organization conducted all steps in the evaluation process; supported evaluations had an evaluator external to the organization involved in conducting some, but not all, steps in the evaluation process, often alongside internal staff; and external evaluations were considered those conducted entirely by external evaluators.

The survey instrument asked respondents to rate their current level of satisfaction with their evaluation processes on a scale from 1 – 5 (not at all satisfied; slightly satisfied; somewhat satisfied; very satisfied; extremely satisfied). The survey also asked for respondents' ratings of the importance of achieving various evaluation goals on a scale from 1 - 4 (not at all important to extremely important) and how well they believe their evaluations have achieved those goals on a scale from 1 - 4 (not at all achieved to entirely achieved). External accountability goals included satisfying grant requirements, using the evaluation results to demonstrate the success of programs to external stakeholders, and learning the extent to which participants are satisfied with their programs. Adaptive management goals included determining if programmatic goals are being achieved, understanding why programs achieve or don't achieve programmatic goals, using the results to make adjustments to improve programs, and understanding how programs affect different audiences. This final goal specifically addresses diversity, equity, and inclusion concerns in the adaptive management process and can help organizations determine how to better serve diverse audiences. Survey items used to measure these concepts can be found in Table 3-2.

Analysis

We analyzed these data using Stata16 software. We report the mean and standard deviations of responses to survey items in Table 3-2. To determine practitioner satisfaction with evaluation processes, we calculated the mean satisfaction score for all participants and ran a one-way ANOVA to determine differences in mean satisfaction across evaluation types (i.e., informal, formal internal, formal supported). We then conducted a multiple linear regression analysis with listwise deletion and ordinary least squares estimation (see Dismuke & Lindrooth, 2006) to understand drivers of practitioner satisfaction. Independent predictor variables included the achievement score of each evaluation goal and evaluation type with informal evaluations designated as the reference group for this categorical variable. Because only five respondents

reported having an external evaluator conduct their entire evaluation, we grouped external evaluations with supported evaluations for these analyses[3].

To identify potential areas for improvement within current evaluation processes, we calculated the mean importance and achievement scores of each evaluation goal as well as the overall mean importance and achievement scores across all evaluation goals, resulting in an importance-performance analysis (see Martilla & James, 1977). Finally, to identify potential approaches or strategies to improve evaluation goal achievement, we conducted a one-way ANOVA to compare mean evaluation goal achievement scores by type of evaluation conducted (i.e., informal, formal internal, formal supported).

Results

Survey respondents

In total, 538 EE practitioners responded to the survey. Respondents indicated that their organizations provide EE programs all over the world, but primarily within the United States (93% of respondents). Figure 3-2 shows the locations of programming provided by respondents' organizations within the United States. A breakdown of respondents' organization type, organization size, funding sources, roles, program types, audiences served, types of evaluations conducted, and personnel who conduct evaluations can be found in Table 3-1. Thirty-eight of these respondents indicated they did not conduct evaluations, and thus were not included in this study. The percentage of respondents who indicated they do not conduct evaluations (4%) is likely an underrepresentation due to the non-random sample we obtained; however, this study is explicitly focused on those who currently conduct evaluations, which makes this underrepresentation less problematic for these analyses.

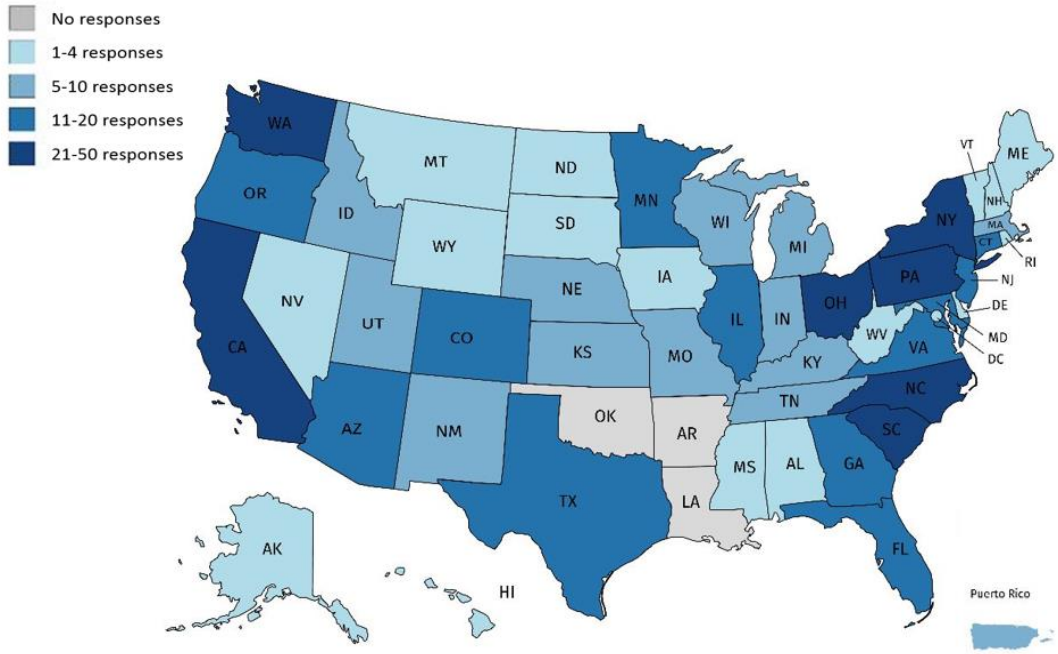

Figure 3-2. Locations of programming provided by respondents' organizations within the United States

[3] The results did not differ when these responses were removed from the data set.

Table 3-1. Summary of respondent, organization, and program characteristics.

Variable	% of respondents
Organization type *(n=518)*	
Science, Nature, or Community Center	19%
College, University, or Research organization	19%
Non-profit organization (not in another category)	18%
National or State Park, Protected area, or Cultural site	11%
K-12 School	11%
Local, State, or Federal Government (not parks)	10%
Aquarium, Museum, Zoo, Farm, or Garden	9%
Camp or Residential Center	3%
Organization size *(n=524)*	
Very small (<10 employees)	41%
Small (10-49 employees)	36%
Medium (50-249 employees)	15%
Large (250+ employees)	8%
Funding sources *(n=375)*	
Foundation or government grants	66%
Membership, program, or visitor fees	64%
Philanthropic donations	45%
Government funding	40%
Fundraising events	35%
Other[1]	17%
Respondent's role *(n=536)*	
Teach EE	82%
Hiring, training, or managing	79%
Conduct or make decisions about program evaluations	62%
Raise funds	32%
Executive decision maker	31%

Variable	% of respondents
Program type *(n=530)*	
School field trips	63%
Visits to a school	45%
Summer camps	44%
Nonformal programs with walk-in visitors	29%
After school programs	18%
Exhibit-focused visits	18%
Multi-day overnight experiences	15%
Other[2]	10%
Audience ages (% who serve this group often) *(n=529)*	
Pre-K (younger than 5 years old)	28%
Grades K-4 (ages 5-10)	65%
Grade 5 (ages 10-11)	67%
Grades 6-8 (ages 11-14)	50%
Grades 9-12 (ages 14-18)	33%
Adults (ages 18+)	48%
Audience identities (% who serve this group often) *(n=523)*	
People of color	50%
People for whom English is not their first language	25%
People from lower socioeconomic groups	58%
Type of evaluations conducted *(n=523)*	
Both formal and informal	57%
Informal only	34%
Formal only	5%
No evaluations are conducted	4%
Personnel conducting formal evaluations *(n=303)*	
Internal	52%
Supported	46%
External	2%

[1]Other funding sources included operating budgets, gift shop and other sales, and corporate sponsorships.
[2]Other program types included virtual programs, events and festivals, classes and workshops, and professional development.

Survey items and descriptive statistics

Overall, mean satisfaction with evaluation processes for all respondents who answered this question (n=447) was 2.69 on a scale from 1 – 5 (not at all satisfied to extremely satisfied). The most important evaluation goals reported by respondents were learning the extent to which participants are satisfied with their programs, determining if programmatic goals are being achieved, and using the evaluation results to make adjustments to improve programs (Table 3-2). The goals that were most highly achieved included satisfying grant requirements, learning the extent to which participants are satisfied with their programs, and using the results to make adjustments to improve programs.

Table 3-2: Survey item descriptions with mean and standard deviations.

	Evaluation goal importance		**Evaluation goal achievement**	
	How important do you consider each of the following potential goals for your organization's EE program evaluations?		To what extent do you feel your evaluations have achieved each goal?	
	1-4 Scale[1]		*1-4 Scale*[2]	
Survey Items	**Mean**	**SD**	**Mean**	**SD**
External accountability goals	3.37	0.57	2.88	0.63
· Satisfy grant requirements	3.02	1.11	2.99	1.02
· Use the results to demonstrate the success of our programs to external stakeholders	3.44	0.75	2.73	0.84
· Learn the extent to which participants are satisfied with our programs	3.65	0.54	2.96	0.7
Adaptive management goals	3.53	0.49	2.51	0.65
· Determine if we are achieving our programmatic goals	3.71	0.52	2.73	0.76
· Understand why our programs achieve (or don't achieve) programmatic goals	3.53	0.7	2.34	0.81
· Understand how our programs affect different audiences.	3.14	0.86	2.06	0.84
· Use the results to make adjustments to our programs.	3.75	0.52	2.87	0.78

[1]Scale labels: 1- not at all important, 2- slightly important, 3- moderately important, 4- extremely important
[2]Scale labels: 1- not at all achieved, 2- somewhat achieved, 3- mostly achieved, 4- entirely achieved

Identifying drivers of practitioner satisfaction

ANOVA results indicate that respondents who conduct formal evaluations, either internally or with external support, had a significantly higher mean satisfaction score compared to respondents who reported only conducting informal evaluations (Table 3-3).

Table 3-3. ANOVA results of evaluation satisfaction score by evaluation type.

	Informal (I) (n=178)	Formal Internal (F) (n=157)	Formal Supported (S) (n=146)	ANOVA f (df) p
Mean satisfaction	$2.32^{F,S}$	2.80^{I}	2.99^{I}	24.45 (2,441) <0.01

Superscript indicates significance at p<0.01. Dunnett's T3 post-hoc test used. Eta-squared[4] = 0.10.

Regression results indicated that achieving adaptive management goals, understanding the extent to which participants were satisfied with their experience, and conducting formal evaluations were correlated with a greater satisfaction with evaluation processes (Table 3-4). This model accounted for 47% of the variance in practitioner satisfaction.

Table 3-4. Summary of regression analysis for predictors of respondents' satisfaction of evaluation processes.

Variable grouping	Independent variables	β	SE	t	p-value
Achievement of External Accountability Goals	Satisfying grant requirements	0.06	0.04	1.42	0.16
	Determining participant satisfaction	0.12	0.06	2.51	0.01
	Demonstrating success to stakeholders	0.04	0.06	0.78	0.44
Achievement of Adaptive Management Goals	Determining if goals are being achieved	0.18	0.07	3.14	<0.01
	Determining why goals are or are not being achieved	0.26	0.07	4.42	<0.01
	Understanding how programs affect different audiences	-0.01	0.05	-0.10	0.92
	Using results to improve programs	0.18	0.06	3.22	<0.01
Evaluation Type (compared to informal)	Formal internal evaluations	0.10	0.09	2.18	0.03
	Formal supported evaluations	0.14	0.09	2.86	<0.01

n= 358; R^2= 0.47.

Identifying areas for improvement

We conducted an important/performance analysis (see Martilla & James, 1977) to determine key areas for potential improvement within practitioners' current evaluation processes. Across all seven evaluation goals, the overall mean evaluation goal importance was a 3.46 and the overall mean evaluation goal achievement was a 2.67 (on a scale from 1-4). These mean values represent the origin of the axes in Figure 3-3. The x-values indicate mean goal importance and the y-values indicate mean goal achievement. Each evaluation goal, as well as the overall means for external accountability goals and adaptive management goals, are represented on the chart. The resulting importance-performance (or achievement) chart displays four distinct quadrants. Quadrants 1 and 2 represent areas that are being achieved to a greater extent than the average goal achievement, while Quadrants 3 and 4 highlight key areas for improvement.

[4] An eta-squared value of 0.01-0.08 indicates a small effect, 0.09-0.24 indicates a medium effect, and 0.25+ indicates a large effect.

- **Quadrant 1** (Q1): Gauging participant satisfaction, determining if programmatic goals are being achieved, and using evaluation results to inform programmatic improvement were considered of higher importance and were also more commonly achieved.
- **Quadrant 2** (Q2): On average, respondents felt they were successful at satisfying grant requirements and using the results to demonstrate success to external stakeholders, although these goals were of lower average importance. Taking the average of the three external accountability goals, external accountabilities fall in this quadrant.
- **Quadrant 3** (Q3): Understanding how programs affect different audiences was considered of lower importance and was also less commonly achieved.
- **Quadrant 4** (Q4): Understanding why programs achieve or don't achieve programmatic goals falls in this quadrant, with above average importance and below average achievement. Taking the average of the four adaptive management goals, adaptive management also falls in this quadrant.

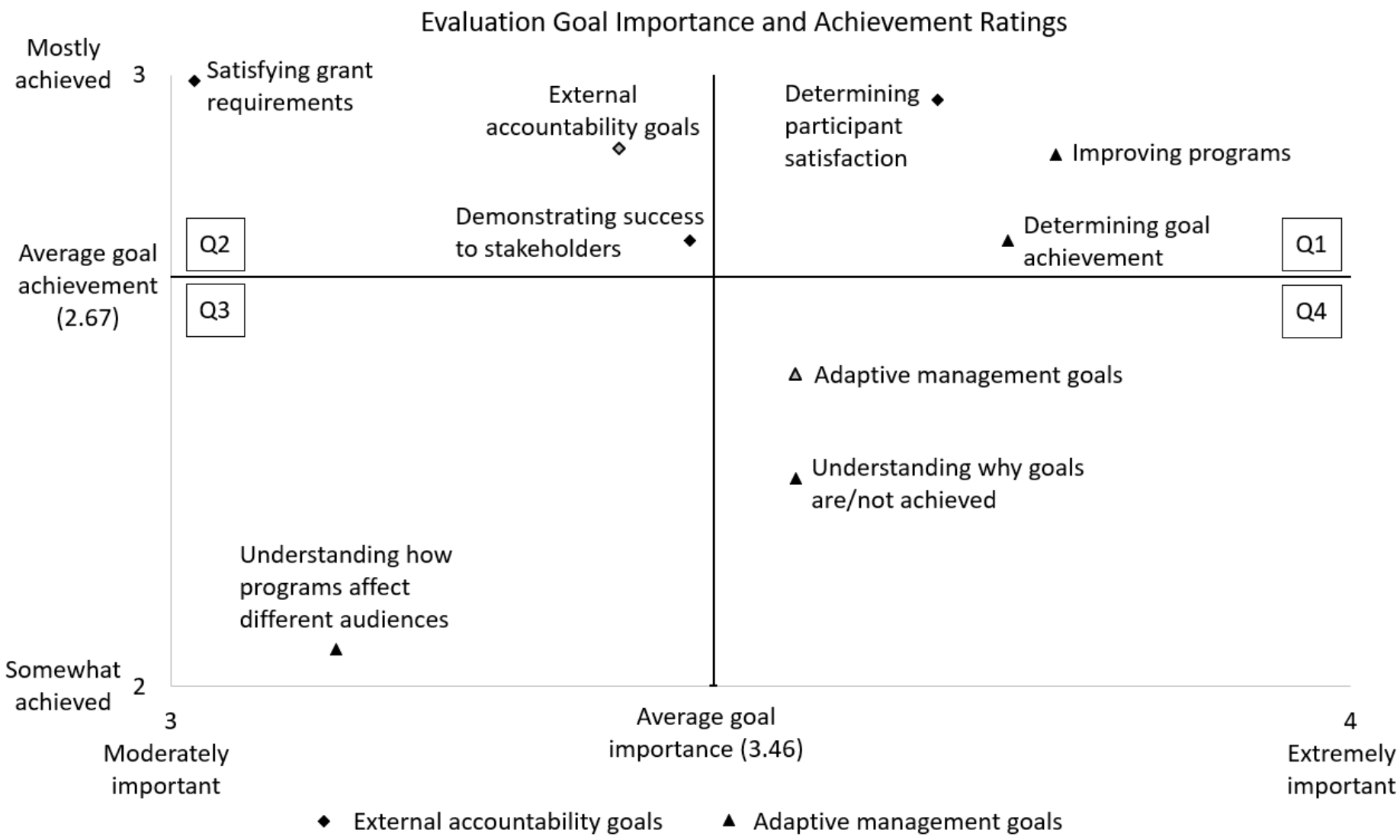

Figure 3-3. Average evaluation goal importance and the extent to which each goal is achieved.

Identifying approaches for improvement

A one-way ANOVA showed that organizations who conducted formal evaluations, either internally or with the support of an external evaluator, achieved evaluation goals to a greater extent compared to those who conduct only informal evaluations (Table 3-5). Additionally, formal supported evaluations outperformed informal evaluations and formal internal evaluations with regard to achieving external accountability goals, specifically satisfying grant requirements and demonstrating success to external stakeholders.

Table 3-5. ANOVA results comparing mean evaluation goal achievement by evaluation type

		Mean achievement score				
		Informal (I)	**Formal Internal** (F)	**Formal Supported** (S)	**ANOVA**	
Goal Type	**Evaluation Goal**	*(n=178)*	*(n=157)*	*(n=146)*	**p**	**f (df)**
External accountability	Satisfying grant requirements	2.83^{S*}	2.88^{S*}	$3.28^{I*,F*}$	<0.01	7.44 (2,381)
	Determining participant satisfaction	$2.80^{F,S*}$	3.01^{I}	3.08^{I*}	<0.01	6.47 (2,424)
	Demonstrating success to stakeholders	$2.41^{F*,S*}$	$2.75^{I*,S*}$	$3.05^{I*,F*}$	<0.01	22.33 (2,417)
Adaptive management	Determining if goals are being achieved	$2.51^{F*,S*}$	2.81^{I*}	2.92^{I*}	<0.01	12.19 (2,427)
	Understanding why goal are or are not being achieved	$2.10^{F*,S*}$	2.41^{I*}	2.55^{I*}	<0.01	12.58 (2,420)
	Understanding how programs affect different audiences	1.95	2.07	2.18	0.08	2.57 (2,410)
	Using results to improve programs	$2.70^{F,S*}$	2.94^{I}	3.01^{I*}	<0.01	6.47 (2,422)

Superscript indicates significance at p<0.05; Superscript* indicates significance at p<0.01.
Dunnett's T3 post-hoc test used to determine significant differences.

Discussion

Understanding and enhancing practitioner satisfaction with their evaluation processes is likely a critical step towards promoting the widespread use of evaluation for programmatic improvement in EE. Our results indicate that practitioner satisfaction with evaluations processes is low, and mean satisfaction did not surpass the midpoint (or somewhat satisfied) on a 1-5 scale. Achieving adaptive management goals, determining participant satisfaction with their program experiences, and using formal evaluation methods tended to enhance practitioner satisfaction with their evaluation processes. While adaptive management goals were considered more important to achieve than external accountability goals, they are being achieved to a lesser extent. Our results highlight two areas for improvement, including understanding and executing the adaptive management process and using evaluation to further DEI efforts in the field of EE. Formal evaluation processes were more likely to achieve these, and all other evaluation goals to a greater extent than informal evaluations and thus identifies a pathway forward for improving practitioner satisfaction and evaluation processes.

Practitioner satisfaction may be low for a number of reasons, including a reliance on informal evaluation processes, a lack of knowledge and capacity to conduct formal evaluations (Powell *et al.* 2006), or an incomplete understanding of how evaluation can be used to make evidence-based programmatic improvements. While about one-third of survey respondents indicated they conduct only informal evaluations, respondents that conducted formal evaluations tended to have higher levels of satisfaction with their evaluation processes. This result is likely influenced by our finding that formal evaluations achieved evaluation goals to a greater extent compared to informal evaluation processes. Formal evaluation consists of forming evaluation questions, creating evaluation tools, designing data collection processes, determining sampling procedures, implementing the evaluation, analyzing and interpreting the data, and reporting on evaluation results (Mertens & Wilson, 2018; Newcomer, Hatry, & Wholey, 2015). Some organizations may have staff with evaluation knowledge and expertise that can implement this process; however, many organizations do not have the capacity, knowledge, time, or resources to conduct a high-quality formal evaluation (Keene & Blumstein, 2010; Powell *et al.* 2006). Partnering with an external evaluator can help to ensure the evaluation process runs smoothly and produces valid and actionable data. External evaluators can also help train organization staff and build evaluative thinking skills. While organizations may not have the capacity to evaluate their programs internally, learning from evaluators about the evaluation process can help practitioners think more intentionally about their intended programmatic goals and how their organization can better achieve those goals.

Respondents who successfully used evaluations to understand participant satisfaction with the program tended to have higher levels of satisfaction with their evaluation processes. This evaluation goal is already being highly achieved, on average, by respondents' current evaluation processes. Using participant satisfaction surveys to understand the extent to which participants were satisfied with their experience is a common use of evaluation in EE (Stern *et al.*, 2014), and may be what many practitioners are familiar with when they think about evaluation (Carleton-Hug & Hug, 2010). This may be due to the ease with which organizations can measure participant satisfaction on a large scale with few resources. Understanding participant satisfaction should not be discounted; however, evaluations have the potential to do much more for an organization than reveal participant satisfaction. Our results show that achieving adaptive management goals may be equally important to enhance practitioner satisfaction with evaluation processes.

Using evaluation results to improve programs was rated, on average, as the most important goal to achieve with an evaluation; however, there may be a lack of understanding about the adaptive management process and how to execute it within the EE community. We measured the achievement of three sequential steps in the adaptive management process, including 1) determining if programs are achieving programmatic goals, 2) understanding why programmatic goals are or are not achieved, and 3) using evaluation results to inform programmatic improvement. While the first and third steps in the adaptive management process were achieved to a greater extent than the average, the second step, understanding why programs meet or do not meet programmatic goals, was achieved to a lesser extent, below the average achievement score. It is unclear how organizations might use evaluation data to improve their programs without an understanding of how various programmatic elements influenced the measured outcomes. This mismatch in the data could stem from an incomplete understanding of

the adaptive management process (Morghan, Sheley, & Svejcar, 2006), the complexities involved in ascertaining what programmatic elements are driving participant outcomes, or a reliance on assumptions made about how and why programs are more or less successful.

Evaluation can be used to determine which programmatic elements are leading to the achievement or lack of achievement of programmatic goals by collecting data on the implementation process of the program (Linnan & Steckler, 2002). Conducting a summative evaluation at the conclusion of a program may help organizations determine if programmatic goals are being met, but without data to capture what led to the achievement of those goals, practitioners will not be able to use that information to make evidence-based adjustments to programming (Fien, Scott, & Tilbury, 2001). To understand what programmatic elements, such as the instructor's delivery, the preparation and information the participants received before arriving, the presence and engagement of teachers or chaperones, or specific pedagogies employed, may lead to more positive participant outcomes, the measured outcomes can be compared across programs with differing characteristics. The best practices identified in this stage of the evaluation can then be implemented with a new group of program participants, and their outcomes can be compared to previous participant outcomes. This will provide evidence as to whether the program adjustments led to the intended change in participant outcomes. When organizations commit to adaptive management, evaluations have a much greater potential to be used to inform programmatic improvement, benefitting the organization and its program participants.

While respondents identified the three adaptive management goals outlined above as relatively important to achieve with evaluations, understanding how programs affect different audiences was considered of least importance and was achieved to a much lesser extent than all other evaluation goals. This may stem from organizations not considering their participant base to be diverse; however, within this sample, 50% of organizations often serve people of color, 25% often serve people for whom English is not their primary language, and 58% often serve people from lower socioeconomic groups. Alternatively, organizations may not see evaluation as a means for learning how to better serve diverse audiences. The literature reveals the many unique barriers people of color can face to participate in outdoor programs, activities, and recreation, including limited access to outdoor areas and programs, limited communication between environmental organizations and communities of color, a fear of discrimination, and a lack of diversity within outdoor and environmental organization staff (Johnson, Bowker, English, & Worthen, 1997; Pease 2015; Roberts, 2007; Warren, Roberts, Breunig, & Alvarez, 2014). Additionally, traditional EE practices, founded primarily on White, Western ideals, may not resonate with all audiences (Finney, 2014; McLean, 2013; Neilson, 2008) and therefor may not achieve the same outcomes as White participants. Evaluations can be a useful tool to further understand an organization's participant base and how different groups are responding to the program (Braun, Cottrell, & Dierkes, 2018; Stufflebeam & Shinkfield, 2012).

Determining how different audiences are impacted by a program requires critical and reflexive thought when both designing an evaluation and analyzing evaluation data (Hall, 2020). There is a growing movement in the evaluation field towards culturally responsive evaluations (Askew, Beverly, & Jay, 2012; Boyce & Chouinard, 2017). This approach entails a collaborative process including a variety of stakeholders, specifically those from marginalized groups and the

populations served, to design evaluations that are sensitive to and valid for the culture and context in which the program is located (Acree & Chouinard, 2020; Askew *et al.*, 2012; Greene, 2006; SenGupta, Hopson, & Thompson-Robinson, 2004). Evaluations that are culturally responsive are more likely to measure salient outcomes for a particular audience using culturally appropriate language. After collecting evaluation data, the data can be disaggregated into groups of interest, whether that be by race, ethnicity, native language, gender, age group, or any other relevant groupings to determine how the program affected each group. By relying on aggregate data and interpreting averaged results, it is impossible to determine if certain groups are benefiting more from the program than others. The evaluation results can then be used to make evidence-based adjustments to programming to meet the needs of all participants.

Evaluation can be a complex, resource-intensive process, and when the end result does not provide tangible, actionable results, practitioners may feel their efforts have been wasted. Adding to this problem, the fear of conducting evaluations that show less than perfect success can lead practitioners to avoid evaluation all together (Carleton-Hug & Hug, 2010). Funders and other stakeholders that require evaluations to demonstrate success can make the consequences of a negative evaluation seem devastating. However, from an adaptive management perspective, failure indicates opportunity for growth and improvement. To increase the use of formal evaluations, enhance practitioner satisfaction with evaluation processes, and improve evaluation processes, the field of EE may need a shift in thinking, from evaluating to determine and demonstrate success, to evaluating with the intent of finding areas for improvement. Reframing evaluation in this way can help EE adapt to the ever-changing social, political, and environmental climate, and remain resilient in the face of uncertainty.

Limitations

The data collected for this research are not necessarily a representative sample of EE practitioners and organizations within the United States or across the globe. Because this survey was administered online via social media and to members of two professional organizations in the field of EE (NAAEE and ANCA), not all EE practitioners had the opportunity to have their voice heard and contribute to these data. These data likely represent practitioners who are more active in their respective professional organization and more engaged on social media. Additionally, while we explicitly asked for a broad range of practitioners to respond, regardless of their experience with evaluation, it is likely that those who took the time to respond to the survey were also more familiar with evaluation and more actively engaged in the evaluation of their organizations' programs. While the results shared are not necessarily representative of all EE practitioners, they may offer insights regarding the possible state of evaluation in the field of EE from the perspective of those most engaged in the process, as well as identifying areas for future research and for the improvement of evaluation processes across the field of EE.

Conclusion

This research offers insights into areas and approaches for improving evaluation processes, evaluation goal achievement, and practitioner satisfaction with evaluation processes. Pressures to meet the needs of external stakeholders can be high due to the possible repercussions of failing to comply with these accountabilities; however, using evaluations as an adaptive management tool can increase practitioner satisfaction with evaluation processes and the use of the evaluation beyond complying with reporting requirements. Currently, there may be

an incomplete understanding of evaluation and adaptive management processes within the EE community. Partnering with external evaluation professionals may help organizations build evaluative thinking skills and execute the formal evaluation process, which we found may be better equipped to achieve various evaluation goals compared to informal evaluations. Using evaluations to further DEI efforts in the field may be an area needing further research and clarification. EE practitioners may not see the connection between using evaluations to understand potential disparities in participant outcomes and working to improve the relevance and inclusivity of their programs for diverse audiences. Reframing evaluation as a means to further improve programming to reach intended goals may incentivize organizations to conduct evaluations more frequently and try new techniques, as opposed to framing evaluation as a means to demonstrate success and efficacy. Organizations that commit to formal evaluations with adaptive management goals may find their evaluations more useful and be more satisfied with their evaluation processes moving forward.

References

Aceves-Bueno, E., Adeleye, A. S., Bradley, D., Brandt, W. T., Callery, P., Feraud, M., Garner, K.L., Gentry, R., Huang, Y., ... & Tague, C. (2015). Citizen science as an approach for overcoming insufficient monitoring and inadequate stakeholder buy-in in adaptive management: criteria and evidence. *Ecosystems*, *18*(3), 493-506.

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.

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

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).

Braun, T., Cottrell, R., & Dierkes, P. (2018). Fostering changes in attitude, knowledge and behavior: demographic variation in environmental education effects. *Environmental Education Research*, *24*(6), 899-920.

Capwell, E. M., Butterfoss, F., & Francisco, V. T. (2000). Why evaluate? *Health Promotion Practice*, *1*(1), 15-20.

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

Cordie, A. M. (2009). *The Development, Implementation, Evaluation, and Revision of an Online Course Entitled Making Environmental Education Relevant for Culturally Diverse Audiences* (Doctoral dissertation). University of Wisconsin-Stevens Point.

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

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.

Dismuke, C., & Lindrooth, R. (2006). Ordinary least squares. *Methods and Designs for Outcomes Research*, *93*(93-104).

Fien, J., Scott, W., & Tilbury, D. (2001). Education and conservation: Lessons from an evaluation. *Environmental Education Research*, *7*(4), 379-395.

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.

Frensley, T. B., Stern, M. J., & Powell, R. B. (2020). Does student enthusiasm equal learning? The mismatch between observed and self-reported student engagement and environmental literacy outcomes in a residential setting. *The Journal of Environmental Education*, *51*(6), 449-461.

Giese, J. L., & Cote, J. A. (2000). Defining consumer satisfaction. *Academy of marketing science review*, *1*(1), 1-22.

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.

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.

Ham, S. H. (2013). *Interpretation: A guide to making a difference on purpose*. Golden, CO: Fulcrum Publishing.

Heimlich, J. E. (2010). Environmental education evaluation: Reinterpreting education as a strategy for meeting mission. *Evaluation and Program planning*, *33*(2), 180-185.

Johnson, C. Y., Bowker, J. M., English, D. B., & Worthen, D. (1997). *Theoretical perspectives of ethnicity and outdoor recreation: a review and synthesis of African-American and European-American participation.* Gen. Tech. Rep. SRS-11. Asheville, NC: US Department of Agriculture, Forest Service, Southern Research Station.

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

Linnan, L., & Steckler, A. (2002). *Process evaluation for public health interventions and research.* San Francisco, CA: Jossey-Bass.

Martilla, J. A., & James, J. C. (1977). Importance-performance analysis. *Journal of marketing*, *41*(1), 77-79.

McDuff, M. (2002). Needs assessment for participatory evaluation of environmental education programs. *Applied Environmental Education and Communication: An International Journal*, *1*(1), 25-36.

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

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

Morghan, K. J. R., Sheley, R. L., & Svejcar, T. J. (2006). Successful adaptive management—the integration of research and management. *Rangeland Ecology & Management*, *59*(2), 216-219.

Neilson, A. L. (2008). *Disrupting Privilege, Identity, and Meaning: A Reflective Dance of Environmental Education*. Rotterdam, The Netherlands: Sense Publishers.

Newcomer, K. E., Hatry, H. P., & Wholey, J. S. (2015). *Handbook of practical program evaluation* (4th ed.). San Francisco, CA: Jossey Bass.

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

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

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

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

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

Stern, M. J., Powell, R. B., & Ardoin, N. M. (2010). Evaluating a constructivist and culturally responsive approach to environmental education for diverse audiences. *The Journal of Environmental Education*, *42*(2), 109-122.

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.

Stufflebeam, D. L., & Shinkfield, A. J. (2012). *Systematic evaluation: A self-instructional guide to theory and practice*. Dordrecht, The Netherlands: Springer.

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.

West, S., Schultz, L., & Bekessy, S. (2016). Rethinking social barriers to effective adaptive management. *Environmental management*, *58*(3), 399-416.

A culturally responsive evaluation framework for environmental education

Abstract

Despite the many challenges inherent to 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 United States and globally. We reviewed the literature on culturally responsive approaches to evaluation and propose a culturally responsive evaluation framework for use in the field of EE. This framework helps EE organizations and professional evaluators consider how issues of race, culture, 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. Finally, we call for organizations that support and fund EE efforts to recognize the importance of this approach, encourage its use, and provide adequate resources for its implementation.

Introduction

Despite the many challenges inherent to 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). A culturally responsive approach to evaluating EE programs can help to addresses both the need for more formal, 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 accounts for 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 for adequate resources for its implementation. We also describe avenues for future research on culturally responsive evaluation and its application in EE.

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 (UNSECO, 1976; 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, & Holthius, 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 which 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.

While the field embraced this shift, as it suited the needs of program providers, 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). 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 impact from the viewpoint and voice of the participants (Patton, 2002). 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 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. 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.

Culturally Responsive Approaches to Evaluation

The field of evaluation at large has identified several evaluation 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.

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 a study 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.

What is Culturally Responsive Evaluation?

CRE is a collaborative approach to evaluation that intentionally engages all 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, Hopson, & Thompson-Robinson, 2004).

Evaluators using this approach employ democratic principles to facilitate deliberations among stakeholder groups to identify relevant evaluation questions, inform culturally appropriate methods of inquiry, and assist in interpreting evaluation data (Acree & Chouinard, 2020; Thomas & Parsons, 2017). Culturally responsive evaluators question normative assumptions based on Western ideals and traditional methodologies that exclude non-dominant epistemologies and fail to consider the lived experiences of program stakeholders (Hall, 2020; SenGupta *et al.*, 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 confounding 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).

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 entire evaluation process. This process involves identifying stakeholders that may be missing from the process, determining relevant evaluation questions that attend to stakeholder needs, defining appropriate evaluation methodologies and data collection procedures, collecting data, interpreting the data, and disseminating evaluation results in meaningful ways to diverse audiences (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.*, 2014; 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, Archiblad, Hargraves, & Trochim, 2015). By being involved in the evaluation process, stakeholders are more likely to understand the evaluation findings and put them to use (O'Sullivan, 2012; Torres *et al.*, 2000).

For EE programs, engaging in CRE may mean expanding traditional views of who is included as stakeholders and sharing power and decision-making responsibilities to determine relevant areas of inquiry, interpret the data, and make recommendations for program improvement. Relevant stakeholders could include program staff from various organizational levels, program participants, parents and chaperones, teachers, school administrators, local community members, and program funders. 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.

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 minutia 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 designing culturally appropriate methodologies, but it also provides insights for interpretating 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 deeply considering the cultural appropriateness of the modes of inquiry; 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.

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 population served 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 research, consult with program participants and other stakeholders to ensure language and communication styles reflect local norms, validate instruments with the target population, and consider the use of non-traditional methods of data collection (Askew *et al.*, 2012; Chouinard & Cousins, 2009). For EE programs, this may mean expanding the breadth of evaluation methods and measures beyond pre/post participant surveys to include mixed-methods and innovative means of data collection that are attuned to the culture of program participants. Pre-testing and pilot testing measures and instruments with the target audience can further test culture appropriateness.

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. Creating and validating survey instruments can be a long and arduous process, with many iterations tested 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 websites such as My Environmental Education

Evaluation Resource Assistant (MEERA; https://meera.snre.umich.edu/) and the North American Association of Environmental Education (NAAEE) Evaluation Portal (https://evaluation.naaee.org/).

Using a standardized measure as part of a program evaluation does not preclude it from being culturally responsive. Rather, 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 other forms of data collection that can supplement survey findings. Collecting qualitative data at each program site can help evaluators further understand the culture of each program and community and to elicit multiple perspectives in a way that reflects the lived experiences and voices of the program participants (Briggs *et al.*, 2019). This 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, facilitating democratic processes, and focusing on culture and context, CRE highlights historically marginalized perspectives that enlighten the evaluation process. Evaluators are encouraged 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 scientific field, 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 from and bearing the costs of the program.

A Culturally Responsive Framework for Evaluating Environmental Education Programs

We now outline a culturally responsive evaluation framework adapted from Frierson, Hood, and Hughes (2002) that provides guidance on how to implement this type of evaluation in the field of EE (Figure 4-1). 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. However, the degree to which an organization takes 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 (or those affected by or that otherwise have an interest in the program) 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 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.

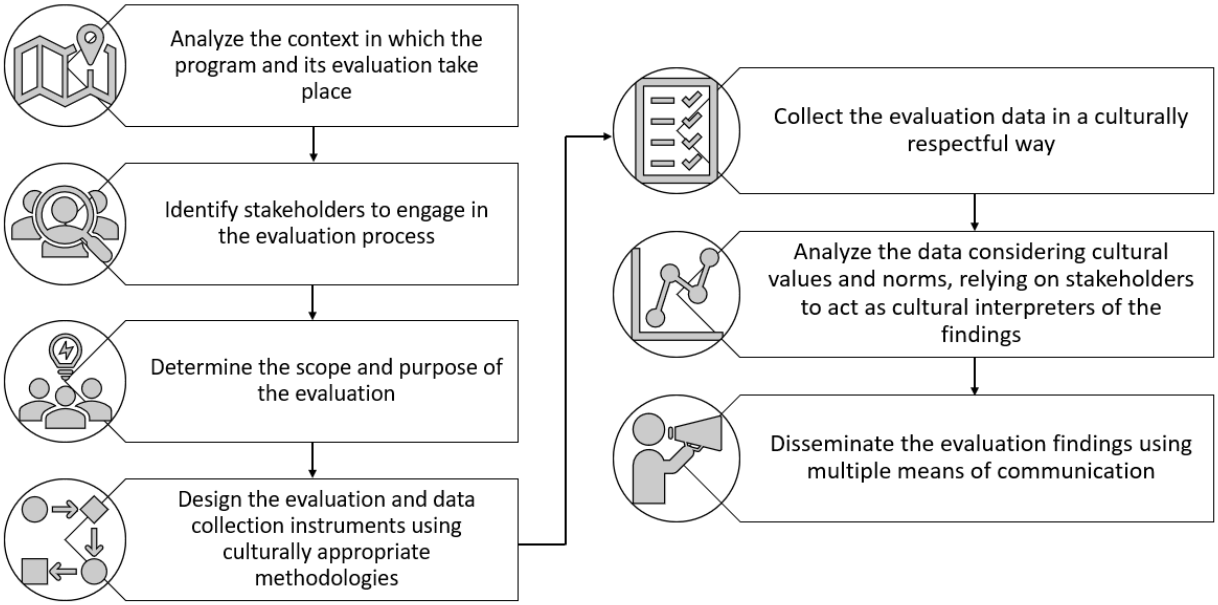

Figure 4-1. A culturally responsive evaluation framework.

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). 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, and historical 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? And 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 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 excluded from participating in the program?
- 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 is included and who is excluded 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).

Critical questions:

- Who is currently involved in the evaluation process and to what extent?
- Are marginalized, underserved, and underrepresented groups adequately represented?
- Who is missing from the evaluation process?
- How can we meaningfully engage stakeholder groups who are not currently or traditionally involved in the evaluation process?
- Are there imbalances of power and privilege among stakeholder groups?
- How might we engage teachers, school administrators, parents, and program participants in the evaluation process while 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 democratic processes to facilitate discussions among stakeholder groups and between these groups and the evaluator. Building trust between stakeholders and in the evaluation process can help to reduce feelings of risk that can stifle healthy collaboration. 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. The first two types reflect trust between individuals; the third reflects trust in the set of formal or informal rules or procedures that govern their interactions. Rational trust involves predicting the likely outcomes of the behaviors of one's counterparts in a collaborative process. Thus, 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 and demonstrations of active caring and listening. 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). Understanding the dynamics of these dimensions of trust among stakeholder groups can help evaluators facilitate collaboration and work to build trust in areas where it is lacking.

To determine the evaluation questions and programmatic outcomes that will guide the evaluation process, evaluators might consider using a deliberative democratic approach (House & Howe, 1999). This approach involves the inclusion of all relevant stakeholder interests when considering potential evaluation questions, dialogue between stakeholder groups to identify shared and divergent interests and concerns, and deliberation between stakeholders and evaluators to reflect on issues raised. To avoid conflict between stakeholders and arrive at agreement, evaluators can use a principled negotiation strategy (Fisher, Ury, & Patton, 2011). This strategy seeks to focus stakeholder attention on the problem or issue to be resolved and not on the emotional tension between parties. This requires setting some clear guidelines for interactions that create a safe space for disagreement. Principled negotiations avoid positional bargaining, or zero-sum bargaining focused on proclaiming specific idea, rather focusing on the interests that underlie these ideas. If the interests of all parties can be identified, these interests can serve to create a list of criteria by which to consider potential pathways forward. The more participants can make the effort to find and express merit in others' ideas, to focus on addressing each other's interests rather than arguing for specific programmatic preferences, and to use mutually agreed upon-criteria for assessing potential pathways forward, the better chances they will have to develop a system that feels valid to all parties (Stern & Baird, 2015; Fisher & Shapiro 2005).

Critical questions:

- Does the evaluation process foster the building of trust between stakeholders, program staff, and evaluators?
- Can stakeholders agree on a process for decision-making that feels fair and safe?
- How might imbalances of power and privilege impact the democratic processes used?
- How can we ensure all relevant stakeholder interests are 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 of determining relevant evaluation questions?

- 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 stipulations. 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. Additionally, evaluators are encouraged to use both quantitative and qualitative methods to collect data from program participants and other stakeholders so that the evidence can be triangulated to depict a holistic picture of the program's value and impact (Stickl Haugen & Chouinard, 2019). Data collection processes should be developed specifically for the target audience 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 instruments with the target audience can help to 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?
- Which of the identified designs and methodologies most closely align with cultural norms?
- Does the sampling scheme adequately represent marginalized or underrepresented groups?
- 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 relationship with external researchers and seek to directly address negative perceptions of research and evaluation. 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. Note, too, that compensation or other meaningful recognition for community members for their role in the evaluation can be considered, especially when community members have previously been subjected to extractive or otherwise oppressive research or evaluation experiences.

Critical questions:

- Who is providing the evaluation data?
- How might the complex history of research with marginalized groups influence participants' willingness to cooperate?
- How can evaluators communicate the value of the evaluation to participants in a culturally meaningful way?
- 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 and norms, relying on stakeholders to act as cultural interpreters 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 multiple evaluators and 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:

- How are different forms of data analyzed 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?
- What contextual factors need to be considered when interpreting the data?
- Which cultural groups are included in the data and how can they be aggregated and disaggregated to understand the nuances between and within groups?
- Do certain groups disproportionately benefit from the program?

7. Disseminate the evaluation findings using multiple means of communication

The evaluator, along with colleagues and collaborators, should share the evaluation findings widely, particularly to stakeholder groups that were engaged in the evaluation process. Program stakeholders are also encouraged to be involved in presenting and dissemination the

evaluation findings. Often, an evaluation report is required by funders, although this report is not the only way in which evaluation findings should be disseminated. Different stakeholder groups may have various understandings of statistical methods and 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).

Critical questions:

- 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?

Call to Action

Conducting an evaluation using a culturally responsive approach is hard work and requires substantial resources. While staff members, directors, and managers internal to an EE organization can make use of this proposed framework and critical guiding questions, this approach is likely best suited for a participatory process involving an external evaluator to serve as both a methodological expert and as a facilitator of the process. 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. 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.

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.

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

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.

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

Belgrade Charter, *Connect*, UNESCO-UNEP Environmental Education Newsletter, 1976(1)

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).

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

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.

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. (2010). Serving the public good. *Evaluation and Program Planning*, *33*(2), 197-200.

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. 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.

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

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., & Howe, K. R. (1999). *Values in evaluation and social research*. Thousand Oaks, CA: Sage Publications.

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.

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. (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.

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.

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.

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.

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

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.

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

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

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.

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.

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.

Chapter 5

Conclusion

Introduction

This research and dissertation were almost entirely conducted and written during COVID-times. This presented unique challenges, having to deviate substantially from my research proposal. As the world came to a screeching halt in March of 2020, so too did environmental education (EE) programs. Students were learning from the comfort of their home as they stared at computer screens. Field trips were no longer being offered and content outside of the required curricula, which tends to include EE, was shed to provide much-needed extra time to cover basic curricula to meet national and state standards. Without students participating in EE programs, I lost many opportunities to collect data from students, hear their perspectives, and understand their experiences. With these set-backs came new opportunities to explore lines of inquiry that were on the forefront of my mind as well as the field of EE at large. This dissertation is the beginning of my journey to tackle three major challenges in the field of EE: 1) increasing attendance to EE programs and addressing the disparities in attendance between racial groups; 2) improving evaluation processes to meet the needs of EE practitioners; and 3) promoting diversity, equity, inclusion (DEI), and cultural responsiveness in the evaluation of EE programs. In this conclusion, I will highlight two cross-cutting themes prevalent throughout this dissertation and carry these themes through as I reflect on my experiences and take a look towards the future.

Cross-cutting themes

Evaluation in Environmental Education

The chapters within this dissertation highlight the utility of evaluation in EE and the need for more formal, systematic evaluations and an increased capacity to implement them within the EE community. For more widespread evaluation to occur, practitioners need to have confidence in and be satisfied with their evaluation processes and understand how evaluations can be used as a tool for organizational learning and programmatic improvement. Chapter 2 gave us insight into what might be driving practitioner satisfaction with evaluation processes and identified areas for potential improvement. Conducting formal systematic evaluations with an explicit focus on adaptive management may move the field in the right direction. Evaluations conducted solely for accountability purposes miss opportunities for learning and growth. Evaluations can be resource intensive and time consuming, and when they do not produce actionable results, the cost can feel greater than the benefits. As formal evaluations become more prevalent and EE organizations build capacity to conduct evaluations or acquire funds to hire evaluation professionals, EE programming will continue to improve and positively impact future generations' ability to combat the wicked sustainability challenges that our world faces.

Chapter 1, although not explicitly pertaining to or referencing evaluation, provides an example of where evaluation could help further DEI efforts in the field of EE. The literature is ripe with evidence pointing to the lack of diversity in the environmental field and the barriers faced by people of color to participate in environmental programs and outdoor recreation. Our results echoed those findings as students of color were less likely to attend NorthBay, even

though NorthBay's programming was explicitly designed for diverse urban audiences. This calls attention to the need for evaluation to understand what aspects of NorthBay's recruitment, outreach, and programming may be influencing the disparities we see in program attendance. Conducting evaluations with an adaptive management mindset, NorthBay could try out new recruitment and outreach strategies, pedagogical techniques, and program structures to determine what practices lead to a diverse participant base that more accurately reflects the racial/ethnic make-up of the areas they strive to serve.

Chapter 3 outlines a framework for designing and implementing evaluations in a culturally responsive way, leading to more valid evaluation results. If evaluations are not conducted in a way that produces valid results, any potential programmatic adjustments informed by those results will not likely be successful. When working in a multicultural space, using traditional Eurocentric evaluation methods may not produce results that are valid for the culture and context in which the program and its audiences are located. Culturally responsive evaluations also provide the opportunity for diverse collaborations and partnerships in which new ideas for measures, methods, and processes can be formed. The sustainability challenges we face will need collaborative efforts like this to provide the knowledge, skills, and motivation to all audiences to combat these global challenges.

Diversity, Equity, and, Inclusion in Environmental Education

Aspects of DEI permeate each chapter of this dissertation. In Chapter 1, we found that students' race had a significant impact on their intentions to attend a 5-day residential EE programs, with students of color being far less likely to plan to attend. The underlying factor of parental support, which also varied significantly by race, is a clear indication of the systemic issues of discrimination that are pervasive across the United States and in the environmental field. As Nina Simon eloquently states in her book *The Art of Relevance*, it is not enough to open the existing door wider. We must co-create new doors that are relevant and enticing to diverse audiences. That is, increasing access to EE programs in the form of subsidizing transportation costs and program fees, intentionally reaching out to underserved audiences, hiring people of color, and providing translators or offering programs in multiple languages, are all strategies that open the existing door wider. These strategies are necessary but insufficient to increase participation in EE from underserved audiences. To create new doors includes co-creating programs from the ground up, working with communities served to understand their needs, their interests, and their stories, creating content that is relevant to and reflects their lived experiences, and implementing programs in ways that incorporate diverse cultures, and ways of knowing and interacting with the world.

Chapter 2 highlights the missed opportunities to use evaluation to understand and address the needs of diverse audiences in EE. When evaluations only consider average impact, it is impossible to determine who is benefiting from the program most, and where disparities might lie. Diversity can take many forms pertaining to any number of the following categories: race/ethnicity, culture, urbanity, age, gender, socioeconomic status, sexual orientation, religion, disability, and more. Chances are nearly every EE program offered serves a diverse audience. Evaluation can help reveal disparities that may exist by highlighting groups that benefit more from the programs than others. Collecting and reflecting on this information is step one, followed by enacting adjustments and changes to program content and delivery to meet the needs of all

participants. Through adaptive management and the co-creation process described above, organizations can use evaluation to improve their programming, creating new doors and opportunities for traditionally underserved audiences to experience, feel safe in, and learn from the environment.

Chapter 3 outlines a culturally responsive approach to evaluation in EE which considers how the issues of race, culture, power, privilege, and inequity influences the evaluation process and the validity of evaluation findings. While this process can be complex and we advise that organizations who wish to undertake a culturally responsive approach to work with an external evaluator, the chapter also provides considerations for how EE practitioners can be more reflexive in their thinking and throughout evaluation process. This collaborative approach is intentional about engaging with all relevant stakeholders, particular those who are marginalized, underserved, or do not traditionally have a 'seat' at the table. This could include students or other program participants, parents of students who participate, classroom teachers, school administrators, community groups, organization staff and leadership, or program funders. This approach requires stakeholder input in all steps of the evaluation processes, including determining the purpose of the evaluation, the evaluation questions, and the programmatic outcomes, designing the evaluation and data collection instruments, collecting the data, analyzing and interpreting the data, and disseminating the evaluation findings. We propose guiding questions that evaluators and program staff can ask themselves as they navigate the evaluation process to ensure issues of race, culture, power, privilege, and inequity are addressed within each step. Using a culturally responsive approach to evaluation supports organizational learning, furthers social justice efforts and produces evaluation findings that are culturally valid.

Reflection

My experience as a Virginia Tech Hokie spans nearly a decade. I first called Virginia Tech home in 2010 as I began my freshman year in the College of Natural Resources and the Environment and now in 2021, I am graduating from my alma mater with a doctorate degree in a field that inspires me every day. During the first two years of my program I had the opportunity to work with an exceptional EE organization, NorthBay, and saw firsthand the impact a program like this can have on young people. For some, this experience was life changing, and for me it opened my eyes to the possibilities of what can be achieved with EE. The final year of my Ph.D. program was spent more isolated, in front of a computer screen in the midst of a global pandemic. I spent several months taking a deep dive into the published literature, which helped me make sense of my experiences. My collective experiences while conducting this research solidified the two main themes that emerged throughout my dissertation work.

Evaluation is complex and resource-intensive

Conducting formal, systematic evaluations require a thorough understanding of the research process, the goals of the organization, and the program logic. Often times, EE organizations lack clarity in all three of these areas. EE organizations tend to operate on tight budgets, limited staff, and expedited timelines. The lack of adequate resources, knowledge, and skills to design an evaluation while simultaneously designing the program itself can lead to programs with inadequate program logic, where organizations rely on experience and assumptions to make decisions about how to structure the program to meet intended goals. In some instances, programs may be designed without specific programmatic goals in mind. While

program staff can rely on experience as to what tends to lead to a successful program, the definition of 'success' may be called into question without an evaluation of previous programs. As noted in Chapter 2, even programmatic outcomes like engagement and enjoyment may be misunderstood by program staff without formally evaluating these outcomes. Designing the evaluation in concert with designing the program can help ensure the program logic aligns with organizational goals for the program, and that these goals are appropriately measured in the evaluation.

An organization that starts the evaluation process without much forethought during the program design phase may come to frustrating realizations. They may determine their programs lack clear programmatic outcomes or goals or that their program logic is not sufficient to meet intended goals. They may also come to determine that their program is not as 'successful' as they previously had thought before conducting an evaluation. These realizations can be devastating to an organization that is passionate about meeting their mission and working to make their community a better place. As evaluators, we must be supportive during these times, highlight successes, and have constructive conversations about areas for improvement. This is where adaptive management is so critical. Without making modifications and re-evaluating to determine if those revisions made a difference, organizations may feel disempowered by their initial evaluation results. Leaving organization staff feeling worse off than they did before the evaluation process began will not likely promote evaluative thinking or an evaluation culture within that organization.

While I cannot provide organizations with more time or financial capital, I can work to build evaluative thinking skills within each organization I work with. Building organizational capacity to conduct their own evaluations takes extensive time and resources, something that many EE organizations do not have. However, building evaluative thinking skills is much more feasible. By evaluative thinking I mean facilitating an understanding of program logic, intended goals, and indicators of success. Program staff do not necessarily need to know the intricacies of survey development, data collection processes, data analysis, and reporting. These skills take years of training to thoroughly understand and implement, and asking this of EE organizations can lead to staff feeling overwhelmed and inadequate. However, if program staff have the knowledge and skills to design a program with clear program logic, focused on meeting clearly identified goals, and a general sense of what success looks like for them, this can propel the field forward in its own right, empowering organizations to feel confident in their programs without having to make assumptions about success. I fear the movement to build evaluation capacity within organizations may disincentivize practitioners to conduct evaluations after realizing what the process entails. Focusing first on building evaluative thinking, organizations may come to further appreciate evaluation and the benefits it provides without asking too much and overwhelming organizations with the expectation that they formally and rigorously evaluate their own programs.

Ignorance is the enemy of progress in promoting diversity, equity, and inclusion in EE

Here, I refer to ignorance as a general lack of understanding. Often, this ignorance stems from considering your own personal lived experience as common, the norm, or in some ways taken for granted. It also stems from a lack of explicitly seeking to understand other points of view. I suffered from this ignorance before my experiences within this Ph.D. program. My lived

experiences may resonate with you, the reader, or may be in stark contrast to your own lived experiences. I am a White, cisgender, heterosexual female who grew up in middle to upper class suburbia. My elementary, middle, and high school years were spent in a community in which my classmates resembled me in almost every way. My undergraduate experience was similar as I spent my time in classes relating to natural resources and the environment, a predominantly White field. It was not until I was thrust into an environment at NorthBay, where my lived experiences were not shared by the attending students, that I started to grapple with what terms like diversity, equity, and inclusion even mean. These terms have become buzzwords, words that are included in formal language about an organization's guiding principles or missions, words that get tossed around as 'ideals' but never fully defined in context.

For people with lived experiences like mine, promoting DEI may mean making the world a better place, but for people of color and other marginalized groups, promoting DEI means finally getting a seat at the table, it means finally be offered opportunities that many others take for granted, it means feeling a sense if efficacy and voice, it means being granted the same rights, the same access, and the same support that many others are granted the moment they are born. Promoting DEI is not just an ideal to make the world a better place, it is a movement to ensure all people can succeed, can contribute, can be heard, can influence, can inspire, can create change, can feel safe, can feel valued; because of their unique culture and identity, not in spite of it.

The first step is recognizing that others may have very different lived experiences than yourself. Phrases like "I don't see color" or "anyone is welcome to attend our programs" further diminishes the lived experiences of marginalized groups. It does not acknowledge the hardships and barriers marginalized groups face to participate in the same activities, to be given the same opportunities, and to achieve the same success that privileged others are granted at birth. The current issues surrounding race, culture, gender identity, and sexual orientation can be uncomfortable to discuss for those who have lived a privileged life, but without asking the difficult questions and listening to the responses of those who have different lived experiences, we will not make progress.

In EE, this may mean asking students to report their self-identified race/ethnicity on a survey to further understand who is attending and who is not, and who is benefiting most from the program and if these benefits are equitably distributed. It may require collaborating with teachers, students, and parents to understand their needs and the unique barriers they face to participate in EE programming. It may mean diverting from traditional practices used for decades and trying new structures, formats, techniques, and pedagogies. We do not have all the answers or strategies to promote DEI in EE successfully, but we do know that many traditional structures, formats, techniques, and pedagogies implemented in EE has, albeit unintentionally, led to disparities in attendance, participation, perceived relevance, and personal impact. We cannot stand by and let these disparities become further entrenched as the norm. We must pave a new path forward that incentivizes diverse collaborations, promotes the implementation of new ideas and new modes of operation, and accepts that failures mean opportunities for learning and growth. The fear of asking tough questions, taking others' perspectives, and failing to succeed must not prevail if we are to transform EE into a space that is diverse, equitable, and inclusive.

Looking to the future

I feel fortunate that my dissertation and my Ph.D. experience as a whole have propelled me on a journey to promote DEI and improve evaluation processes and education programs in the field of EE, and that this journey has only just begun. It is an intimidating feeling to know you have only scratched the surface of what can be accomplished to further efforts in these areas, but this realization also provides inspiration and the drive to keep going. While I do not know what the future holds or what career path I may take, I know DEI and evaluation will be front and center in my work. Ideally, I would like to continue conducting research and share my findings broadly with the EE community and make steady progress, improving EE, and instilling the skills and motivation necessary for future generations to tackle the wicked sustainability and social justice issues we face today. Below, I outline ideas for future work and lines of research.

Evaluation in Environmental Education

- Develop processes and tools for promoting evaluative thinking within EE organizations that are both feasible and realistic in terms of the knowledge, skills, and resources available to many organizations.
- Conduct case studies of culturally responsive evaluations in practice to refine the framework presented within Chapter 3 of this dissertation.
- Explore the utility of embedded assessments and qualitative approaches to evaluation that may be more easily implemented by EE practitioners.
- Facilitate and refine a collaborative evidence-based learning network process to share knowledge and improve EE programs on a wider scale.

Diversity, Equity, and Inclusion in Environmental Education

- Understand what practices are currently being used to promote DEI efforts in the field and to what extent those practices are meeting intended goals.
- Work with diverse collaborators to identify additional strategies and practices that might promote DEI in the field of EE and evaluate the success of these practices.
- Understand how participant perceptions of program relevance and inclusivity impact student outcomes.
- Conduct qualitative work with students from marginalized groups to understand their perceptions of environmental education and the barriers to attending and participating in EE experiences.
- Understand trends in participation of diverse audiences in various types of EE, including single-day programs, field trips, experiences at museums, zoos, aquariums, and nature centers, residential camps, and virtual programs.
- Develop an effective co-creation process that EE organizations can implement with stakeholders to promote culturally responsive practices and increase program relevance for diverse audiences.

Wherever I end up, I will always center equity, justice, integrity, collaboration, and continuous learning in my work thanks to my experiences at Virginia Tech within my Ph.D. program.

Appendix

Institutional Review Board Approvals

VIRGINIA TECH.

Division of Scholarly Integrity and
Research Compliance
Institutional Review Board
North End Center, Suite 4120 (MC 0497)
300 Turner Street NW
Blacksburg, Virginia 24061
540/231-3732
irb@vt.edu
http://www.research.vt.edu/sirc/hrpp

MEMORANDUM

DATE:	May 29, 2020
TO:	Marc J Stern, Kelley Anderson
FROM:	Virginia Tech Institutional Review Board (FWA00000572, expires October 29, 2024)
PROTOCOL TITLE:	Environmental Education Program Evaluation Constraints
IRB NUMBER:	20-276

Effective May 29, 2020, the Virginia Tech Human Research Protection Program (HRPP) and Institutional Review Board (IRB) determined that this protocol meets the criteria for exemption from IRB review under 45 CFR 46.104(d) category(ies) 2(ii).

Ongoing IRB review and approval by this organization is not required. This determination applies only to the activities described in the IRB submission and does not apply should any changes be made. If changes are made and there are questions about whether these activities impact the exempt determination, please submit a new request to the IRB for a determination.

This exempt determination does not apply to any collaborating institution(s). The Virginia Tech HRPP and IRB cannot provide an exemption that overrides the jurisdiction of a local IRB or other institutional mechanism for determining exemptions.

All investigators (listed above) are required to comply with the researcher requirements outlined at:

https://secure.research.vt.edu/external/irb/responsibilities.htm

(Please review responsibilities before beginning your research.)

PROTOCOL INFORMATION:

Determined As:	**Exempt, under 45 CFR 46.104(d) category(ies) 2(ii)**
Protocol Determination Date:	**May 29, 2020**

ASSOCIATED FUNDING:

The table on the following page indicates whether grant proposals are related to this protocol, and which of the listed proposals, if any, have been compared to this protocol, if required.

Invent the Future

VIRGINIA POLYTECHNIC INSTITUTE AND STATE UNIVERSITY
An equal opportunity, affirmative action institution

Global Leader | Proven Expert

May 7, 2019

Marc J. Stern, PhD, BS, MS
Virginia Tech
310 West Campus Drive
304 Cheatham Hall
Blacksburg, VA 24061

Dear Dr. Stern:

SUBJECT: IRB EXEMPTION—NOT HUMAN SUBJECTS RESEARCH
Investigator: Marc J. Stern, PhD, BS, MS
Institution Protocol #: 19-274
Protocol Title: Barriers preventing students from attending NorthBay

This letter is in response to your request for an opinion as to whether the above referenced research would constitute human subject research requiring IRB review.

This opinion is based on federal regulation 45 CFR 46 and associated guidance.

We determined this study is exempt from IRB review because it does not meet the definition of *human subject research* as defined in 45 CFR 46.102. Specifically this is not a research study designed to develop or contribute to generalizable knowledge. The primary goal of this research is to understand the barriers that are preventing students from attending NorthBay Adventure Center for a three-day residential environmental education program, relevant to students attending Title 1 middle schools in Prince George's County, Maryland.

This determination can apply to multiple sites, but it does not apply to any institution that has an institutional policy of requiring an entity other than WIRB (such as an internal IRB) to make such determinations. WIRB cannot provide a determination that overrides the jurisdiction of a local IRB or other institutional mechanism for making such determinations. You are responsible for ensuring that each site to which this determination applies can and will accept WIRB's determination.

Western Institutional Review Board
1019 39th Avenue SE Suite 120 | Puyallup, WA 98374-2115
Office: (360) 252-2500 | Fax: (360) 252-2498 | www.wirb.com

Please note that any future changes to the project may affect its status as research that does not involve human subjects, and you may want to contact WIRB about the effect these changes may have on the status before implementing them. WIRB does not impose an expiration date on its determinations of research that does not involve human subjects.

If you have any questions, or if we can be of further assistance, please contact Bridget D. Brave, JD, or e-mail RegulatoryAffairs@wirb.com.

BDB:jca
NHS Exemption-Stern (05-07-2019)
cc: VA Tech, WIRB, Virginia Polytechnic Institute and State University
Kelley C. Anderson, Virginia Tech
WIRB Accounting
WIRB Work Order #1-1182147-1