Which approaches are associated with better outcomes? Evidence from a national study of environmental education field trip programs for adolescent youth in the United States

Robert B. Powell, Marc J. Stern & Brandon Troy Frensley

Routledge
Taylor & Francis Group

Which approaches are associated with better outcomes? Evidence from a national study of environmental education field trip programs for adolescent youth in the United States

Robert B. Powell[a] (ID), Marc J. Stern[b] (ID) and Brandon Troy Frensley[c] (ID)

[a]Department of Parks, Recreation and Tourism Management, Department of Forestry and Environmental Conservation, Clemson University, Clemson, SC, USA; [b]Department of Forest Resources and Environmental Conservation, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA; [c]Department of Environmental Sciences, University of North Carolina Wilmington, Wilmington, NC, USA

ABSTRACT

Which approaches are associated with better student learning outcomes in environmental education (EE)? We observed a sample of 299 day-long EE field trip programs occurring across the U.S.A. for youth in grades 5–8 (ages 9 to 14). We tracked the extent of use and quality of implementation of 66 programmatic, educator, and setting characteristics and measured student outcomes immediately after the programs using a retrospective survey. A series of complementary tests identified 11 characteristics that were most powerfully and consistently associated with learning outcomes, accounting for 18% of variance in learning outcomes. These included group size, naturalness, novelty, place-based pedagogy, verbal engagement, quality questions, transitions, and staging, as well as the responsiveness, comfort and clarity, and emotional support provided by the educator. Some of the most commonly promoted practices in the EE field were rarely observed. Implications are discussed for both practice and research.

Introduction

Which programmatic, educator, and setting characteristics are associated with better student learning outcomes in environmental education (EE)? Reviews of the last decade-plus of research in EE reveal that a large proportion of empirical studies have focused on summative evaluations of single programs in both EE and the larger informal science education field (NRC 2009; Ardoin, Biedenweg, and O'Connor 2015; Ardoin et al. 2018; Rickinson 2001; Stern, Powell, and Hill 2014). These studies have measured the influence of programs on multiple outcomes of interest and documented numerous positive influences on student learning and well-being. However, summative evaluations cannot identify the programmatic elements that are most responsible for reported outcomes. Empirical evidence of this nature can only emerge from either experimental or comparative work, where programs are the unit of analysis and different program characteristics are monitored and measured and linked with measured participant outcomes.

Unfortunately, a dearth of this type of study exists in both EE and the broader informal science education field (NRC 2009; Stern, Powell, and Hill 2014). So despite strong evidence that EE programs can achieve positive learning outcomes, research rarely focuses on which approaches are most consistently linked to better programmatic outcomes (Ardoin, Biedenweg, and O'Connor 2015; Bourke, Buskist, and Herron 2014; Stern, Powell, and Hill 2014).

To address this gap, we observed a sample of 299 day-long EE field trip programs occurring across the U.S.A. for youth grades 5–8 (ages 9 to 14) and tracked the extent of use and quality of implementation of 66 programmatic, educator, and setting characteristics. We measured a range of student outcomes immediately after the programs using a retrospective survey. This approach allowed us to begin to answer the following research question: Which characteristics and approaches are associated with better participant outcomes (i.e. learning, interest in learning, 21st century skills, meaning/self-identity, self-efficacy, place connection, environmental attitudes, environmental stewardship, collaboration, and school motivation)?

Environmental education field trip programs

In the U.S., it is estimated that thousands of youth attend EE field trip programs with their schools annually (Thompson and Houseal 2020). EE field trip programs involve leaving the school grounds and traveling (Ardoin et al. 2018; Storksdieck 2006). These programs generally aim to enhance awareness and knowledge about the environment and its associated challenges as well as develop the skills, dispositions, and expertise to make informed decisions and take actions to address these challenges (e.g. Ardoin et al. 2018; UNESCO 1977). EE field trip programs also typically seek to meet educational standards (e.g. Powell et al. 2019), inspire place connection (Ardoin 2006; Gruenewald 2003), and improve social and emotional learning (Bowers et al. 2010; Garst, Browne, and Bialeschki 2011; Lerner et al. 2005). For this study, we developed and measured cross-cutting outcomes that reflect the broad goals of EE (10 subscales-See Powell et al. 2019 for more detail) and focused on students in grades 5–8 (ages 9–14), because research suggests that middle childhood is a period of rapid development of higher levels of moral reasoning (Kellert 2002; Eisenberg et al. 1987; Kohlberg 1971) and greater logical and abstract cognitive abilities (Kellert 2002; Piaget 1936). Thus, this age period is considered important for developing the higher-level skills needed to foster environmental literacy and develop a connection with nature (e.g. Kahn and Kellert 2002; Sobel 2008).

Although the content of EE programs can vary, a hallmark of EE is to provide immersive and direct contact with nature through hands-on and engaging techniques and approaches (e.g. Storksdieck 2006; North American Associate of Environmental Education (NAAEE) 2020). Many of the most promoted techniques and approaches are described in NAAEE's Guidelines for Excellence in Environmental Education (NAAEE 2017, 2020) and other recent studies and reviews (Stern, Powell, and Hill 2014; Sobel 2008; Krasny 2020). These characteristics can be organized into three broad categories: programmatic and pedagogical approaches; educator characteristics and practices; and setting characteristics.

Programmatic characteristics

Programmatic characteristics refer to how a program is structured and the types of activities undertaken. These techniques include specific pedagogies, such as investigation-focused, inquiry-based, issue-based, experiential, or place-based education (Woodhouse and Knapp 2000; Jose, Patrick, and Moseley 2017; Moseley et al. 2020; Gruenewald 2008; Stern, Powell, and Hill 2014; Stern and Powell 2020; NAAEE 2017, 2020). Other programmatic approaches reflect crosscutting effective communication practices, such as having an introduction, a clear theme or take-home message, transitions between programmatic elements, and a conclusion that provides opportunities for participants to

reflect upon the take home message and their EE experience (e.g. Ham 1992, 2013; Stern and Powell 2013). Table 2 in the results section provides a description of 43 programmatic characteristics hypothesized to influence learning outcomes during EE field trips for adolescent youth.

Educator characteristics

Educator characteristics refer to attributes of the educator thought to enhance outcomes, as well as specific behaviors and ways environmental educators interact with their students to create the instructional environment. Attributes such as passion for the topic and program (Tilden 2009), sincere and authentic interactions with participants (Stern and Powell 2013), confidence, and apparent knowledge, all have the potential to impact outcomes in live educational programs (Stern and Powell 2013; Powell and Stern 2013). Providing emotional support to students and being responsive to participant needs have also been shown to enhance learning outcomes in formal classrooms (Hamre and Pianta 2005; Merritt et al. 2012; Pianta, La Paro, and Hamre 2008; Reyes et al. 2012; Rudasill, Gallagher, and White 2010) and more recently in informal and EE settings (e.g. O'Hare et al. 2020). Table 3 in the results section lists and describes 18 educator characteristics and behaviors hypothesized to influence learning outcomes during EE field trips for adolescent youth. We developed two indexes that characterize seven of the specifically observed characteristics (see *Measurement*), resulting in 13 variables to consider in relationship to outcomes.

Setting characteristics

Research also suggests that the educational environment, or the setting in which an informal or non-formal educational program takes place, may meaningfully influence participant outcomes (e.g. Kellert 2005; Powell et al. 2009; Archer and Wearing 2003). For example, the level of naturalness of a setting is thought to enhance mental and physical well-being (e.g. Kuo, Barnes, and Jordan 2019; Ryan et al. 2010) as well as enhance cognition and learning (Born et al. 2001; Kuo, Barnes, and Jordan 2019; Wells 2000; Wells and Evans 2003). The novelty of the setting, or the perception of something new, unique, or unfamiliar (Garst 2018), can inspire curiosity, learning, and collaborative and collective action (Dale et al. 2020; DeWitt and Storksdieck 2008; de Waal 2008; Keltner et al. 2014). The degree of beauty of the setting can also enhance creativity and imagination (Holton 1988), awareness of balance, symmetry, harmony and grace (e.g. Kellert 2008), motivation to participate in science (Chandrasekhar 1987), and connection to place (Gruenewald 2003, 2008). Similarly, the extent to which participants are physically immersed in a natural setting also appears theoretically important (e.g. Garst 2018; Kellert 2002, 2005). Table 4 in the results section describes five setting characteristics hypothesized to influence learning outcomes during field trips for adolescent youth.

Methods

Overview

This study investigated EE field trip programs for adolescent youth (grades 5–8; ages 10–14) in the U. S. provided by 90 program providers in 24 states to examine 1) which programmatic, educator, and setting characteristics were used during these programs and 2) the relationship between programmatic, educator, and setting characteristics and student outcomes, which were assessed by surveying participating students immediately after the programs. Clemson University's Institutional Review Board reviewed all procedures described below prior to data collection and determined that procedures were Exempt. Verbal informed consent was received from program providers, educators, and students prior to all data collection procedures per guidelines outlined under approved protocols (IRB00000481; FWA00004497).

Selection of sites

Working with the North American Association of Environmental Education (NAAEE), the National Park Service (NPS), and the Association of Nature Center Administrators (ANCA), we identified over 300 potential organizations, including nature centers, botanical gardens, science museums, and national, state, and local parks, that appeared to offer single-day EE field trip programs for students in grades 5–8 (ages 10–14) across the United States during the period of research (i.e. January–June 2018).

We aimed to develop a sample that would represent a broad range of field trip programs in a wide variety of contexts that would be generally reflective of field trip programs across the United States. To do so, we relied on Ruggiero's (2016) evaluation of Environmental Literacy Plans in the US, which ranked states in terms of the status and quality of their statewide Environmental Literacy Plans, as a proxy for the general status of EE in each state. We divided the states into quartiles based on this evaluation and then systematically sought to observe programs of at least 10 program providers from states in each quartile (see Dale et al. 2020 for more information). We then sought to maximize diversity in terms of both program types and socioeconomic context. Inclusion criteria for this study included: programs took place away from school; programs focused on EE (broadly defined); programs lasted a single day or less in duration; programs served grades 5–8; program providers expressed a willingness to participate, and program providers conducted multiple programs during the period of research.

After contacting each potential program provider, we identified clusters of willing providers in different regions of the country that offered single day field trip programs for youth in grades 5–8. Ultimately, we observed 345 programs provided by 90 unique organizations across 24 states and the District of Columbia: 18 providers from the first quartile, 39 providers from the second quartile, 19 providers from the third quartile, and 14 providers from the fourth quartile.

Pilot testing

We developed and refined our observational techniques and data collection procedures based on prior research in formal education (e.g. Gage and Needels 1989; Pianta, La Paro, and Hamre 2008; Stronge, Ward, and Grant 2007, 2011), on interpretive programs in U.S. National Parks (Stern and Powell 2013), and at a residential EE center in Maryland, USA (Frensley, Stern, and Powell 2020; Frensley et al. 2022). We then undertook extensive pilot testing with the entire research team, comprised of eight field researchers and the principal investigators of the study. The research team observed multiple filmed programs followed by 17 in-person field trip programs for the target student population in diverse contexts. During these pilot studies, all members of the research team scored each program as individuals and then compared and discussed at length all discrepancies in scoring. This process was used iteratively to clarify and refine the operational definitions and/or measurement of each programmatic, educator, and setting characteristic under consideration. We used this process to develop consistent (within and across researchers), reliable (produces the same observational result under the same conditions within and across researchers), and valid (accurately measures the construct in question) scoring of all observations across the eight field researchers.

Data collection

Four pairs of researchers visited and collected data at 345 EE field trip programs for 5th to 8th graders. A single pair of researchers visited each location for observation and data collection. Once onsite, the researchers introduced themselves at the outset of each program and then systematically monitored and recorded the extent and quality of programmatic, educator and setting characteristics on a predesigned observation sheet.

For the first two weeks of program observation, each pair of researchers observed programs together and completed initial scoring independently. Afterwards, each pair of researchers would

discuss their observations and scores, which enabled each team to reach consensus on the measure of each indicator, ensuring reliability and consistency in scoring of observational variables. After roughly two weeks for each pair, discrepancies in scoring were rare. Researchers then began to occasionally observe programs individually at the same location. Throughout the 22-week field season, researchers periodically observed programs together again to ensure reliability and consistency in scoring of each variable. All team members (i.e. four pairs of researchers, PIs, and Co-PI) also met periodically online to clarify any questions about scoring. At three staggered points in time over the course of the study, the original pairs of researchers were purposefully intermingled to observe programs together to further enhance the reliability of observation measures.

Immediately following each program, all attending students were invited to complete a paper survey regarding their opinions of the program and its influence on them, which was used to assess the outcomes of the program. For all programs, we attempted a census of all eligible attendees. No time limit was given for the students to complete the survey. The average completion time was approximately 8 minutes. Overall, 5,317 surveys were collected from participants from 345 programs, and the average response rate was 81%.

Data cleaning procedures

Data from the 5,317 surveys were entered into Microsoft Excel and then transferred to SPSS for screening and analysis. First, we dropped three programs (26 surveys) because response rates were below 50%. We then screened surveys for missing values and removed 210 surveys that were missing responses to more than 25% of the items. With these removals, one additional program dropped below a 50% response rate. It was removed entirely (8 additional surveys). We also screened for obvious patterns indicating invalid responses, such as no variability in answers, strings of consecutive numbers, or using one circle to indicate responses for multiple survey items. We identified and removed 94 surveys with these problems. One additional program dropped below 50% response rate following these removals. It was removed from the database along with seven additional surveys. Data were then screened for multivariate outliers using Mahalanobis Distance (MAH) (Tabachnick and Fidell 2018). A total of 563 respondents were removed for exceeding the criterion Mahalanobis Distance value. Six more programs dropped below 50% response rate and as a result were removed from the database (dropping an additional 33 surveys). The resulting sample contained 4,376 individual surveys from 334 programs provided by 90 organizations in 24 states and Washington, DC.

Measurement

Outcomes. Conducting a large-scale comparative study required a consistent tool for measuring outcomes across a wide range of programs that would be valid, reliable, and sensitive enough to vary depending on the quality of the programs. The process of developing this tool included 1) reviewing the literature, 2) involving stakeholders and program providers in workshops to define and refine crosscutting outcomes applicable to a range of EE field trip programs; 3) operationalizing the outcomes following recommended scale development procedures (e.g. DeVellis 2003), which included iterative stakeholder review to ensure external validity; and 4) conducting a series of pilot studies in diverse EE settings across the US to refine and cross-validate the scales using confirmatory factor analyses and multi-group invariance testing procedures (see Powell et al. 2019 for full description). This work resulted in a scale comprised of ten related outcomes, including Place Connection, Learning, Interest in Learning, 21st Century Skills, Meaning/Self-Identity, Self-Efficacy, Environmental Attitudes,

Environmental Stewardship, Collaboration, and School Motivation (see Table 1). All outcomes were measured with multiple items that were measured on a scale of 0–10. Self-Efficacy and Environmental Attitudes were measured using retrospective pre/post questions asking students to reflect on how they felt about given statements before the program, and after as a result of the experience. The mean scores reported for these items represent the differences between pre and post scores. The overall outcome measure, EE21 (short for Environmental Education Outcomes for the 21st Century), is a single composite measure representing the mean of each subscale equally weighted and aggregated to the program level (see results Table 1). The EE21 composite index has been statistically validated through confirmatory factor analysis (Powell et al. 2019).

Because our research seeks to identify which programmatic approaches and characteristics positively influence EE21 irrespective of age and race, and because previous research (e.g. Browning and Rigolon 2019; Browning and Locke 2020) and our own analyses (Stern, Powell, and Frensley 2022) have demonstrated that grade level and race can have strong influences on learning outcomes, we controlled for their influence by group mean centering the EE21 scores (Tabachnick and Fidell 2018). This procedure resulted in us dropping 27 programs from the sample because participants were enrolled in multiple grades. We eliminated another eight programs from analysis because of missing data pertaining to race, which was based on both self-report from students and school data pertaining to racial majority (see Stern, Powell, and Frensley 2022). This resulted in a final sample of 299 programs included in these analyses.

Programmatic, educator and setting characteristics. At each EE field trip program, we recorded observations of 66 different programmatic/pedagogical, educator, and setting characteristics (see results Tables 2–4). Fifty-five constructs were scored on a 1-to-4 scale, two constructs on a 1-to-3 scale, and seven constructs with binary measurement (presence/absence); two were continuous variables. The 1-to-4 scales followed the logic of calibration, as discussed by Ragin (2008) and used in prior research (e.g. Stronge, Ward, and Grant 2007, 2011; Stern and Powell 2013; Frensley et al. 2022): 1 represented total absence; 2 represented minor presence; 3 represented moderate presence; and 4 represented that the characteristic was a dominant aspect of the program. Extensive pilot testing with the full research team revealed that these scales enabled easy categorization, especially for more ambiguous cases, by considering whether the observed program *more* or *less* reflected the characteristic in question (the difference between a 2 and 3 on the scale). It also maximized scale size, which is desirable for detecting meaningful differences between programs and their characteristics. Three-point scales were used for constructs with less variability. One three-point scale measured staging, or the general state of the group upon arrival to the field trip. In this case, the scale represented (3) well-organized; (2) moderately organized (some chaos or confusion); (1) disorganized, frenzied, late, or generally negative. In the other case, the three-point scale represented the extent to which the educator advocated for a specific viewpoint or action. In this case, the scale reflected (3) advocacy was clearly present; (2) vague or implied advocacy, unclear; and (1) not present. Binary variables indicate the simple presence of absence of a characteristic (see results Tables 2 and 3).

Based on our prior analyses (O'Hare et al. 2020; Stern and Powell 2013) and theory (e.g. Pianta and Hamre 2009), we conducted exploratory factor analyses and reliability analyses on the educator characteristics prior to further analyses. We did not conduct

confirmatory factor in this case because educator and program characteristics are formative variables that were observed and represent a specific practice or attribute that is thought to directly influence a dependent variable. This is opposed to reflective indicators, which are thought to represent a broader concept and are not directly observed (see Kline 2005; Jarvis et al. 2003; Podsakoff et al. 2003, for further explanation). Exploratory factor analyses and reliability analyses on program level data revealed the presence of two latent educator characteristics. We have named the two resulting educator factors 'clarity and comfort' and 'emotional support' leaving us with 61 variables to carry forward in subsequent analyses (see results Table 3).

Analyses

To describe the programmatic, educator, and setting characteristics of our sample, we conducted frequency and central tendency analyses. Eleven variables demonstrated little to no variability and were thus dropped from further analysis (see results Table 2). This resulted in 50 characteristics included in our final analyses. To answer our research question, we conducted bivariate Pearson correlation analyses to explore linear relationships between each of the programmatic, educator, and setting characteristics and group-mean-centered (GMC) EE21. We conducted independent samples t-tests on the binary observational items to assess if their presence or absence was related to student learning outcomes. To account for the potential of Type 1 errors, we used a Bonferroni correction to identify the strongest statistically significant relationships (in this case, $p \leq .001$ for 50 distinct analyses).

To further examine the characteristics associated with better learning outcomes, we compared programmatic, educator, and setting characteristics of the top performing quartile of programs versus the lowest performing quartile using independent samples t-tests and chi-squared tests. To account for the potential of Type 1 error, we again used a Bonferroni correction ($p \leq .001$). We also computed Cohen's d effect sizes for the t-tests, which assess the meaningfulness each statistically significant difference between groups (Tabachnick and Fidell 2018). Cohen's d scores below 0.2 can be considered spurious, above 0.2 are considered small, those approaching 0.5 demonstrate a medium effect size, and those nearing or above 0.8 are considered as large effect sizes (Cohen 1992). We calculated phi as the effect size for chi-squared tests. Phi scores below 0.1 are negligible; up to 0.3 are small; between 0.3 and 0.5 are medium; and 0.5 and greater are large (Tabachnick and Fidell 2018).

Finally, we took the programmatic, educator, and setting characteristics that were significantly correlated at the Bonferroni correction threshold and conducted relative weight analysis (RWA) using RWA-Web (Tonidandel and LeBreton 2015), a form of dominance analysis, to examine their relative weight or influence on GMC EE21. RWA iteratively assesses the contribution of an independent variable (in this case, programmatic, educator, and setting characteristics) to the prediction of a dependent variable (in this case, GMC EE21) by itself and in combination with other predictor variables (Budescu 1993; Budescu and Azen 2004; Tonidandel and LeBreton 2011, 2015; Tonidandel, LeBreton, and Johnson 2009). This analysis is warranted when there are large numbers of independent variables that are correlated with each other (Nathans, Oswald, and Nimon 2012; Tonidandel and LeBreton 2011) and provides the relative strength or 'dominance' of an independent variable, as compared with the other independent variables across all potential combinations (Budescu 1993; Budescu and Azen 2004; Tonidandel and LeBreton 2011, 2014). We conducted RWA because many of the 11 predictor variables were correlated with each other, and because RWA provides not only relative weights, but also 95% confidence intervals for the individual relative weights (Johnson 2004) and corresponding significance tests based on bootstrapping with 10,000 iterations (Tonidandel, LeBreton, and Johnson 2009). The RWA also provides a measure of the overall variance in GMC EE21 explained by the dominant predictor variables.

Results

Sample description: programs

All descriptive statistics reported are calculated from the 299 programs validated by data cleaning procedures and group-mean centering of the EE21 dependent variable. Forty-three percent of these programs served 5th graders, 32% served 6th graders, 19% served 7th graders, and 6% served 8th graders. Of these programs, 46% were composed of a majority of students who identified as White, 32% were composed of a majority of students who identified as Hispanic, 13% of programs had no racial majority, and 9% of programs were composed of a majority of students who identified as Black. Free and reduced lunch statistics were available for 275 of the visiting school groups. The proportion of students eligible for free or reduced price lunches ranged from 2% to 100%, with a mean of 57%, similar to the national average of 58% in 2018 (USDA Food and Nutrition Service 2020).

The mean program time was 186.6 minutes, with a standard deviation of 69.9 minutes. The mean group size was 16 students with a standard deviation of 7.3. This number reflects the group size that participated in a program with a unique educator or educators. Many times school groups were broken up into smaller and more manageable sub-groups for programs with a primary instructor. Visiting school teachers were passive observers in 43% of programs. In the other 57%, they played a more active role by participating along with students, providing disciplinary support, and/or co-leading programs. The on-site environmental educators at each location were also asked prior to their programs to identify their desired outcome goals for participants: 68% sought to influence knowledge, 10% skills, 25% interest in learning, 34% attitudes, 34% place attachment, and 24% desired to influence behaviors.

Descriptive statistics: outcomes (EE21)

Table 1 displays the mean and standard deviation for each item that composes EE21 as well as the grand mean and standard deviation for the EE21 composite scale from the 299 programs prior to group mean centering to control for the influence of grade and race (Table 1).

Use of programmatic, educator, and setting characteristics

Programmatic characteristics:. To examine which techniques and approaches were used across the 299 EE field trip programs, we report the mean score and the frequency of use of the different program, educator, and setting characteristics. For 11 characteristics, one-half of the scale (2 points on the observation scale) contained less than 5% of the observations, indicating that we either nearly always or never observed that construct's presence across the sample (Table 2). Due to their low variance, we removed them from further analysis. The programmatic approaches that were rarely observed included: individual and group reflection, group discussion, presenting multiple view-points, teaching 21st Century Skills, role playing, storytelling, and the experiential learning cycle. Other results of note include that observed programs were commonly primarily focused on conveying factual information (approximately 78% were moderate-to-dominantly fact-focused), provided few opportunities for free exploration (15% of programs provided moderate or high levels of free exploration), omitted a clear theme/message (34% provided a clear theme or message), and often lacked a conclusion (just over 50% had a conclusion). Approximately 55% of programs were largely place-based.

Educator characteristics. The majority of educators associated with these programs were highly responsive (83% scored a 3 or 4), were knowledgeable and comfortable

Table 1. EE21 mean and standard deviation of items.

Constructs and items	M (SD)
Place connection: The development of appreciation for and positive personal relationships with the physical location.	
Knowing this place exists makes me feel good.	7.58 (1.26)
I want to visit this place again.	7.57 (1.50)
I care about this place.	7.97 (1.21)
Learning: Knowledge regarding the interconnectedness and interdependence between human and environmental systems	
How different parts of the environment interact with each other.	7.08 (1.10)
How people can change the environment.	7.48 (1.22)
How changes in the environment can impact my life.	7.54 (1.13)
How my actions affect the environment.	7.85 (1.13)
Interest in learning: Enhanced curiosity, increased interest in learning about science and the environment	
Science.	6.50 (1.52)
How to research things I am curious about.	6.58 (1.51)
Learning about new subjects in school.	6.27 (1.56)
21st Century skills: Critical thinking and problem solving, communication, and collaboration	
Solving problems.	5.76 (1.57)
Using science to answer a question.	6.38 (1.48)
Listening to other people's points of view.	6.76 (1.47)
Knowing how to do research	6.48 (1.61)
Meaning/Self identity: A heightened sense of self-awareness, critical reflection, and purpose.	
Taught me something **that will be useful to me** in my future.	6.83 (1.44)
Really made me think.	6.88 (1.47)
Made me realize something I never imagined before.	6.64 (1.54)
Made me think differently about the choices I make in my life.	6.75 (1.53)
Made me curious about something.	6.82 (1.36)
a**Self-efficacy (Retrospective pre-post):** Belief in one's own ability to achieve one's goals and influence their environment.	
I believe in myself.	0.91 (0.75)
I feel confident I can achieve my goals	0.84 (0.60)
I can make a difference in my community.	1.16 (0.65)
a**Environmental attitudes (Retrospective pre-post):** Sensitivity, concern, and positive dispositions towards the environment.	
I feel it is important to take good care of the environment.	0.79 (0.51)
Humans are a part of nature, not separate from it.	1.03 (0.64)
I have the power to protect the environment.	1.22 (0.68)
Environmental Stewardship: Motivations to perform stewardship-related behaviors.	
Help to protect the environment.	7.57 (1.24)
Spend more time outside.	7.30 (1.23)
Make a positive difference in my community.	7.24 (1.26)
Collaboration: Motivation to collaborate more with others.	
Listen more to other people's points of view.	6.99 (1.34)
Cooperate more with my classmates.	6.97 (1.35)
School motivations: Motivation to work harder in school.	
Work harder in school.	7.29 (1.48)
Pay more attention in class.	7.25 (1.54)
EE21 Composite before Group Mean Centering; Cronbach's Alpha=.964	5.84 (1.02)

aRetrospective pre-post items represent the difference between what students' reported feeling after the program and what they felt before.

leading the programs (over 94% scored a 3 or 4), provided clear instructions and content delivery (over 86% scored a 3 or 4), and provided substantial emotional support (Table 3).

Setting characteristics. Programs occurred mostly outside (84% scored a 3 or 4) with over half occurring in natural (60% scored a 3 or 4) and aesthetically pleasing settings (55% scored a 3 or 4). The results also suggest that approximately 70% of programs provided relatively low levels of physical immersion in the environment (Table 4).

Table 2. Program characteristics means frequencies, and relationships (Pearson's r or t-tests) with GMC EE21.

Program characteristics	Definition	M (SD)	1 Not present	2 Minor presence	3 Moderate presence	4 Dominant aspect	Pearson r with EE21
Group Size (Stern and Powell, 2013; Powell, et al., 2009)	Size of group engaged in program.	16.00 (7.3)					−.258***
Place-based (e.g. Gruenewald 2008)	Degree to which the program emphasized and utilized the unique attributes of the place/resource in the program.	2.63 (.87)	8.7%	36.8%	37.5%	17.1%	.226***
Verbal engagement (Knudson et al., 2003; Moscardo 1999)	Degree to which the educator asked questions or engaged students in dialogue	2.62 (.57)	1.3%	38.8%	56.9%	3.0%	.204***
Question quality (Ham 1992; NPS 2019; Frensley et al., 2022)	Degree to which the educator's questions encouraged critical thinking, problem solving, discussion, or consideration of complexity. A quality question cannot be answered with Yes or No (open ended); and are inviting, inclusive, and non-judgmental. They are often designed to elicit: evaluation, focused attention, inference, problem-solving, drawing comparisons, cause-and-effect, personal experiences, and application of new knowledge to new setting.	2.18 (.71)	16.1%	52.2%	29.8%	2.0%	.203***
Transitions (e.g. Frensley et al., 2022; Ham 1992; Jacobson 1999)	Degree to which educator incorporated transitions during shifts in location or content to keep the students engaged and did not detract from the lesson's sequence (*quality not quantity focused*).	2.11 (.76)	21.5%	48.0%	28.2%	2.3%	.196***
Staging (Wong and Wong 2005)	The general tone or organizational state of the group upon arrival and immediately **prior to the program.**	2.62 (.63)	7.7% *Poor*	22.5% *Moderately organized*	69.8% *Well-organized*		.191***
Lecture-based (Stern & Powell, 2013)	The degree to which a traditional lecture approach was used. Traditional lecture may include Q&A and PowerPoint or other visual aids, but it is clearly didactic in nature (one-way communication).	2.14 (.71)	16.1%	57.2%	23.4%	3.3%	−.179**
Class management (Frensley et al., 2022; Pianta and Hamre 2009)	Degree to which the educator effectively managed disruptions or off-task behavior by being proactive and redirecting individuals	3.05 (.86)	4.7%	20.4%	39.8%	35.1%	.177**
Conclusion (e.g. Ham 1992)	Degree to which the educator provided a clear conclusion that included a takeaway lesson or meaningful reflection.	1.85 (.94)	46.0%	29.2%	18.5%	6.4%	.157**
Physical comfort of students	Degree to which the audience was physically comfortable throughout the program	2.61 (.56)	3.3%	32.8%	63.5%	0.3%	−.127*
Free exploration (Bamberger and Tal 2007; Frensley et al., 2022)	Degree to which students were encouraged to explore the environment in their own way	1.72 (.72)	43.1%	42.5%	13.7%	0.7%	.123*
Teacher/chaperone engagement (Fenichel and Schweingruber 2010; NRC 2009; Stern et al., 2008)	Degree to which the educator purposefully engaged visiting teachers or chaperones in the program.	1.41 (.68)	68.4%	23.9%	6.1%	1.7%	.122*
Physical exertion Duration (Minutes)	Degree of physical exertion demanded by program. Total minutes on site	1.73 (.66) 186.6 (69.9)	37.8%	53.5%	7.0%	1.7%	.119* .115*
Facts-only focus (Frauman and Norman 2003; Jacobson 1999; Stern & Powell, 2013)	The extent to which the content that was conveyed was merely factual (devoid of deeper thought, consideration or emotions, values, or other non-factual considerations).	2.93 (.71)	3.7%	18.1%	59.9%	18.4%	−.115*
Hands-on (Frensley et al., 2022; NAAEE 2020)	Degree to which students were actively involved (e.g. touching or interacting with nature, props, etc.) during the lesson; not just passive receivers of verbal or visual communication.	2.46 (.81)	9.4%	46.2%	33.8%	10.7%	.104
Time management (Frensley et al., 2022)	The degree to which the students were able to complete all intended parts of the lesson in an organized and cohesive way.	3.59 (.65)	0.7%	7.0%	25.1%	67.2%	.103

(Continued)

Table 2. Continued.

Program characteristics	Definition	M (SD)	1 Not present	2 Minor presence	3 Moderate presence	4 Dominant aspect	Pearson r with EE21
Inquiry-based (NAAEE 2020)	Degree to which the program uses students' own questions or answers to questions to guide the experience. Typically, an instructor will prime students to pose and answer their own questions. Students' questions and answers appear to guide how the program unfolds.	1.40 (.62)	66.2%	28.1%	5.0%	0.7%	.080
Group work (Stern, Powell and Hill 2014)	Degree to which the program requires/explicitly asks students to work with others.	2.07 (.94)	31.1%	40.8%	18.1%	10.0%	.074
Play/game based (Stern, Powell and Hill 2014; Frensley et al., 2022)	Degree to which the lesson actively engages students in games or competition as an intentional teaching technique.	1.38 (.69)	70.9%	22.4%	4.0%	2.7%	−.073
Investigation (NAAEE 2020)	Degree to which the lesson mirrors the scientific process of formulating a research question, developing and testing hypotheses, and drawing conclusions. May be field-based, lab-based, or class based.	1.85 (1.12)	56.2%	18.4%	10.0%	15.4%	.071
Introduction (e.g. Ham 1992)	Degree to which the introduction captured attention and oriented students' to the program and its content.	2.73 (.81)	7.5%	27.5%	49.8%	15.3%	.058
Relevance (NPS 2014; NAAEE 2017)	Degree to which the lesson content references or makes explicit connections to the students' experience outside the realm of the instruction.	1.99 (.62)	19.1%	63.9%	16.4%	0.7%	−.033
Affective messaging (Jacobson 1999; Lewis 2005; Stern, Powell and Hill 2014)	Degree to which the lesson communicated emotions within the content of the program.	1.60 (.65)	48.8%	42.8%	8.0%	0.3%	.025
Advocacy (NAAEE 2017)	Instructor/program was clearly favoring a specific viewpoint policy, or action as better than another (not as socially acceptable behavior) i.e. picking up trash, brushing teeth, recycling	.39 (.67)	71.9% *None*	17.4% *Vague, implied*	10.7% *Clear*		.024
Message (e.g. Ham 1992, 2013)	There is a clearly communicated central message(s)/theme	2.04 (.97)	37.2%	28.7%	26.7%	7.4%	.018
Social-ecological systems (Merritt et al., 2022)	Degree to which the educator makes explicit efforts at demonstrating social-ecological connectivity.	2.10 (.74)	18.7%	56.2%	21.4%	3.7%	−.017
Issue-based (Frensley et al., 2022; NAAEE 2017)	Degree to which the lesson focuses on real-world environmental problems/issues, their consequences, and potential solutions.	1.80 (.66)	32.4%	56.2%	10.4%	1.0%	−.009
Multi-Sensory (NRC 2009; Beck and Cable 2002; Knudson, Cable, and Beck 2003; Lewis 2005; Moscardo1999; Tilden 1957)	Degree to which students were encouraged to explicitly use their senses **in a unique way** to explore a concept, solve a problem, or connect to their environment.	1.60 (.59)	45.5%	49.5%	5.0%	0.00%	−.001
Individual reflection (Stern, Powell and Hill 2014; Frensley et al., 2022)	Degree to which the students are required to explicitly reflect on their learning experiences as an individual—e.g. through journaling or other reflective techniques during the program	1.19 (.46)	83.6%	14.0%	2.0%	0.3%	
Group reflection (Stern, Powell and Hill 2014; Frensley et al., 2022)	Degree to which the students explicitly reflected on their learning experiences as a group (*through discussion or other reflective group work*). This is not merely a summary, but rather provocation to think through lessons learned (meta-cognition).	1.20 (.43)	80.6%	18.4%	1.0%	0.0%	
Discussion (Stern, Powell and Hill 2014; NAAEE 2017)	Degree to which the educator promoted active student discussion instead of the passive receiving of information or didactic exchange of question and answer. This goes beyond student-teacher Q&A toward active discussion **between multiple members of the group.**	1.23 (.45)	77.9%	21.1%	1.0%	0.00%	
Multiple viewpoints (NAAEE 2017; Stern, Powell and Hill 2014)	Degree to which the program explicitly acknowledged multiple perspectives or uncertainty *(often in controversial issues).*	1.13 (.38)	88.6%	9.7%	1.7%	0.00%	

(Continued)

Table 2. Continued.

Program characteristics	Definition	M (SD)	1 Not present	2 Minor presence	3 Moderate presence	4 Dominant aspect	Pearson r with EE21
21st century skills (Fenichel and Schweingruber 2010; Institute of Museum and Library Services 2009)	Degree to which the lessons explicitly aimed to teach skills, especially skill associated with developing environmental literacy, not just using them. Teaching skills involves discussing and reflecting on their use, not just giving directions.	1.16 (.42)	86.0%	12.7%	1.0%	0.3%	
Role play	Students take on specific roles or pretend to be characters within the lesson.	1.25 (.48)	76.9%	21.1%	2.0%	0.00%	
Storytelling (Beck and Cable 2002; Larsen 2003)	Degree to which storytelling was used as an intentional teaching technique	1.23 (.50)	80.4%	16.4%	3.3%	0.00%	
Unrelated content	The extent to which the program includes activities (e.g. random games, etc.) or content that is unrelated to the lesson.	1.28 (.52)	75.6%	21.4%	2.7%	0.3%	
Visibility	The extent to which the students can clearly see demonstrations, props, visual aids, etc.	3.49 (.57)	0.0%	3.7%	43.8%	52.5%	
Appropriate logistics (Jacobson 1999; Knudson, Cable, and Beck 2003)	Degree to which basic audience and program needs were met (e.g. restrooms, weather, technology, accessibility, shade, sun in eyes, etc.)	3.78 (.48)	0.7%	0.7%	19.1%	79.6%	

Binary characteristics	**Definition**	**Not Present (GMC_EE21 mean)**	**Present (GMC_EE21 mean)**	**t**	**df**	**p**
Pace (Jacobson 1999)	Too fast	93% (0.00)	7% (0.04)	2.35	297	.814
	Too slow	78% (0.08)	22% (−0.03)	−3.09	297	.002**
	Just right	30% (−0.19)	70% (0.08)	2.76	297	.008***
Experiential learning cycle (Kolb 1984; Stern & Powell 2020)	Program completed Kolb's experiential learning cycle, comprised of: (1) a concrete experience; (2): reflection; and (3) using the new knowledge in a new application	97%	3%			

*$p \leq .05$.
**$p \leq .01$.
***$p \leq .001$ (Bonferroni Correction threshold).
Items with little variance at either end of the scale (% of Category 3 + 4 or 1 + 2 ≤ 5% of all observations) were dropped from subsequent analysis.

Table 3. Educator Characteristics Mean (Standard Deviation), Frequencies, and Correlation with GMC EE21 (Group mean centered by Grade and Race to control for their influence).

Educator characteristics	Definition	M (SD)	1 Not present	2 Minor presence	3 Often present	4 Dominant aspect	Pearson r with EE21
Responsiveness (Planta and Hamre 2009; Stern and Powell, 2013)	The extent to which the educator responds to student requests, questions, or other cues.	3.10 (.69)	1.7%	14.7%	55.9%	27.8%	.237***
Instructional Clarity (r=.158*): The degree to which instructions are clearly communicated throughout the program.		3.33 (.66)	0.0%	10.4%	46.0%	43.6%	
Comfort and Clarity Index	Content Clarity (r=.165*): The degree to which the educator effectively communicated and translated the content (science) of the program. It was clear and age-appropriate in terms of vocabulary and delivery.	3.19 (.66)	0.3%	13.1%	54.2%	32.3%	.224***
M=3.32 (.48) Cronbach's Alpha=.757	Apparent Knowledge (r=.149*): Degree to which the educator appears to know the information presented in the lesson, able to answer relevant questions, and demonstrates personal knowledge of the resource in which the program takes place (Ham and Weiler 2002; Lewis 2005; Stern and Powell, 2013)	3.37 (.58)	0.0%	4.8%	53.1%	42.2%	
	Educator Comfort (r=.204*): Degree to which the educator appears comfortable with the students without apparent signs of nervousness or self-doubt. (Stern and Powell, 2013; Lewis 2005; Moscardo 1999; Ward and Wilkinson 2006)	3.38 (.59)	0.0%	5.7%	50.2%	44.1%	
	Positive communications (r=.201*): The extent to which the educator provided positive communications to reassure students and establish a psychologically safe environment free from excessive judgment (Planta and Hamre 2009).	2.92 (.66)	2.0%	19.7%	62.5%	15.7%	
Emotional Support Index (Planta and Hamre 2009; O'Hare, et al., 2020) M=2.95 (.58) Cronbach's Alpha=.823	Affinity Seeking (r=.189*): Degree educators checked in to make sure students are OK, and exhibited behaviors such as listening, sensitivity, nonverbal immediacy (smiling, frequent eye contact, forward leans, etc.), altruism, and trustworthiness (Planta and Hamre 2009; Stern and Powell, 2013).	2.68 (.72)	3.0%	38.1%	47.2%	11.7%	.220***
	Sincerity (r=.178*): Degree to which the educator portrays authentic emotion and seems genuinely invested in the messages he or she is communicating as opposed to reciting information (Ham 2009; Stern and Powell, 2013).	3.25 (.67)	0.0%	12.7%	49.5%	37.8%	
Formality (Planta and Hamre 2009)	The degree to which the educator was formal, official, or rigid vs. casual and laid back about the presentation and discipline.	2.71 (.57)	0.7%	33.1%	60.9%	5.4%	-.145*
Audibility	The extent to which the educator can clearly be heard by the audience.	3.40 (.64)	0.00%	8.0%	43.5%	48.5%	.112
Passion (Ham and Weiler 2002; Moscardo 1999; Powell and Stern 2013)	The educator's apparent level of enthusiasm for the lesson content and the site	2.97 (.63)	0.7%	19.1%	62.9%	17.4%	.100
Humor quantity	The extent to which the educator attempts to use humor, sarcasm, or jokes to share the topic with the students.	1.78 (.58)	29.8%	62.5%	7.4%	0.3%	.075
Eloquence (Lewis 2005)	Extent to which the educator spoke clearly and articulately.	2.99 (.36)	0.00%	7.0%	87.0%	6.0%	.070
Personal sharing (Planta and Hamre 2009; Jacobson 1999)	The degree to which the educator shared personal insights or experiences, answered questions about themselves for the audience, or provided their own opinion on topics or events relevant to the program	1.66 (.61)	41.1%	51.8%	7.0%	0.00%	-.047
Humor quality (Ham and Weiler 2002; Knapp and Yang 2002)	How funny is the educator overall? Does the group react positively to the educator's use of humor and seem to enjoy it?	1.68 (.59)	37.6%	56.8%	5.3%	0.4%	.046
Binary characteristics	Definition	Not Present (GMC_EE21 mean)	Present (GMC_EE21 mean)	t	df	p	
Educator identity (Ham 1992; Ham and Weiler 2002; Knapp and Yang 2002; Larsen 2003;	Friend: Educator is outwardly friendly, casual, approachable, mingles informally, uses familiar slang, wants students to like them, likes to keep things fun, etc.	-0.08 (.78)/ 62%	0.13 (.84)/ 38%	2.18	297	.030*	
Wallace and Gaudry 2002)	Authority figure: Educator emphasizes their own role as educator (a.k.a. boss) and focuses on rules and/or authority to communicate.	0.008 (.82)/ 76%	-0.02 (.77)/ 24%	-0.29	297	.771	
	Walking encyclopedia: Educator focused on conveying a large amount of facts (sometimes which are irrelevant), and often uses jargon, etc.	0.07 (.82) 74%	-0.20 (.76) 26%	-2.62	297	.009**	

*p ≤ .05.
**p ≤ .01.
***p ≤ .001 (Bonferroni Correction threshold).

Table 4. Setting characteristics mean (standard deviation), frequencies, and correlation with EE21.

Setting characteristics	Definition	M(SD)	1 Not present	2 Minor presence	3 Often present	4 Dominant aspect	Pearson r with EE21
Naturalness (e.g. Kuo, Barnes, and Jordan 2019)	Degree to which the program takes place in a manmade vs. natural setting	2.64 (.76)	6.7%	33.1%	49.8%	10.4%	.255***
Novelty (e.g. Garst 2018)	Degree to which the setting is unique or special for the audience. In these situations, the students may communicate that the setting is unexpected or unfamiliar and they exhibit heightened focus on the environment.	2.52 (.71)	3.3%	50.8%	36.8%	9.0%	.237***
Immersion (Beck and Cable 2002; Knudson, Cable, and Beck 2003; Lewis 2005; Moscardo 1999)	Degree to which students are physically immersed in the natural environment (muddy, wet, digging in the dirt, etc.)	2.23 (.79)	13.7%	57.2%	21.1%	8.0%	.139*
Beauty (Kellert 2005; Kaplan, Kaplan, and Ryan 1998)	Degree to which the setting is aesthetically pleasing. At the extreme positive end these are amazing, of overwhelming attraction, or mesmerizing that create a "wow" effect in students.	2.62 (.67)	1.7%	43.8%	45.5%	9.0%	.117
Outdoors (e.g. Kuo, Barnes, and Jordan 2019)	Proportion of time spent indoors vs. outdoors	3.27 (.78)	2.3%	13.4%	38.8%	45.5%	.085

*$p \leq .05$.
**$p \leq .01$.
***$p \leq .001$ (Bonferroni Correction threshold).

Programmatic, educator, and setting characteristics and their relationship with learning outcomes

To determine the relationship between each of the programmatic, educator, and setting characteristics and GMC EE21, we conducted bivariate Pearson correlation analyses. Additionally, we conducted independent samples t-tests on the binary observational items to assess if their presence or absence contributed to student learning outcomes. To account for the potential of Type 1 error, we used a Bonferroni correction to identify significant relationships (in this case $p \leq .001$) in all analyses. While 15 programmatic characteristics were statistically correlated with EE21 ($p \leq .05$), only six met the Bonferroni correction threshold (Table 5). These included group size (smaller groups are associated with better outcomes), place-based programs, higher levels of verbal engagement, high quality questions that led to provocation, the use of transitions between program elements, and the organization of the group upon arrival (staging). Three educator characteristics passed the Bonferroni correction threshold ($p \leq .001$) (Table 5). These included high levels of responsiveness to students needs and inquiries; the comfort and clarity index, which includes clear delivery of instructions and program content as well as the educator's apparent confidence and comfort leading the program; and the emotional support index, which includes providing sincere, positive, and reassuring communications and affinity seeking behaviors. Two setting characteristics, naturalness and the novelty of the setting, were significantly correlated with EE21 at the Bonferroni correction threshold ($p \leq .001$) (Table 5). None of the binary observational variables met the Bonferroni correction threshold ($p \leq .001$) although overly slow-paced programs performed significantly worse than other programs (Table 2), and educators that were 'walking encyclopedias' performed worse than other educators ($p = .009$) (Table 3).

Which programmatic, educator, and setting characteristics distinguish the top-performing from the bottom-performing programs?

To explore whether there are distinguishing characteristics and approaches between top and bottom performing programs, we compared the programmatic, educator and setting characteristics associated with the highest performing quartile of programs vs. the lowest performing quartile. We first divided the 299 programs into quartiles based on their group mean centered EE21 score. Tables 6–9 share all statistically significant differences.

The highest performing quartile of programs had significantly smaller group sizes, a greater degree of place-based programming, and higher levels of verbal engagement than the lowest performing quartile of programs (Table 6). The use of transitions, lecture-based (negative), class management, and providing a conclusion were significantly different with medium effect sizes, although these variables did not pass the Bonferroni correction (Table 6). Slow-paced programs were also more common in the lowest performing quartile ($p < .001$), although the effect size was small in this respect (Table 7).

Educators associated with the highest performing quartile of programs were significantly more responsive, demonstrated higher levels of comfort and clarity, and provided greater emotional support than educators associated with the lowest performing quartile (Table 8). Moreover, programs with walking-encyclopedia-style educators were more commonly in the lowest performing quartile ($p < .001$) although the effect size is small (Table 7). The settings associated with the highest performing programs were significantly more natural and novel than the settings associated with the lowest performing quartile of programs (Table 9).

Relative weight analysis

To understand the proportion of the total variance in GMC EE21 explained by the 11 programmatic, educator, and setting characteristics (Table 5) that were most powerfully related to GMC EE21

Table 5. Correlation Matrix of 11 program, educator, and setting characteristics that met Bonneferoni Correction Threshold with EE21.

Variable	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.
1. GMC-EE21	1	-.258***	.255***	.237***	.237***	.226***	.224***	.220***	.204***	.203***	.196***
2. Group Size	-.258***	1	-.255***	-.258**	-.149**	-.207**	-0.039	-0.065	-.224**	-.132*	-.149**
3. Naturalness	.255***	-.255***	1	.474**	0.111	.368**	.166**	.150**	0.089	.181**	.253**
4. Novelty	.237***	-.258**	.474***	1	0.069	.450**	.129*	.194**	.144*	0.111	.216**
5. Responsiveness	.237***	-.149**	0.111	0.069	1	.244**	.535**	.477**	.348**	.371**	.284**
6. Place-based	.226***	-.207**	.368***	.450***	.244***	1	.299**	.335**	.315**	.319**	.376**
7. Comfort and Clarity Index	.224***	-0.039	.166**	.129*	.535***	.299***	1	.518**	.191**	.373**	.412**
8. Emotional Support Index	.220***	-0.065	.150**	.194**	.477***	.335***	.518***	1	.364**	.384**	.436**
9. Verbal engagement	.204***	-.224***	0.089	.144*	.348***	.315***	.191***	.364***	1	.399**	.380**
10. Question quality	.203***	-.132*	.181**	0.111	.371***	.319***	.373***	.384***	.399***	1	.476**
11. Transitions	.196***	-.149**	.253***	.216***	.284***	.376***	.412***	.436***	.380***	.476***	1
12. Staging	.191***	-0.051	.270***	0.014	.299***	.120*	.226***	.178**	0.07	.242***	.155**

* $p \leq .05$.
** $p \leq .01$.
*** $p \leq .001$ (Bonferroni Correction threshold for 50 variables).

Table 6. Significant results of independent samples t-test comparing programmatic characteristic observational scores between top and bottom quartiles.

Programmatic characteristic	Top quartile M (SD)	Bottom quartile M (SD)	t	df	p	Cohen's d
Place-based	3.09 (.79)	2.43 (.89)	4.784	147	<.001***	.79
Group size	13.49 (6.06)	18.92 (9.28)	4.231	147	<.001***	.70
Verbal engagement	2.84 (.47)	2.46 (.67)	4.045	147	<.001***	.67
Transitions	2.29 (.67)	1.89 (.83)	3.257	146	0.001+**	.54
Class management	3.32 (.74)	2.86 (.96)	3.256	147	0.001+**	.54
Lecture	1.97 (.49)	2.32 (.85)	−3.101	147	0.002**	.51
Full day conclusion	2.09 (.98)	1.66 (.73)	3.071	146	0.003**	.51
Question quality	2.39 (.61)	2.07 (.75)	2.854	147	0.005**	.47
Staging	2.77 (.46)	2.53 (.71)	2.49	146	0.014*	.41
Free exploration	1.95 (.71)	1.69 (.79)	2.083	147	0.039*	.34
Educator comfort	2.51 (.65)	2.70 (.54)	2.008	147	0.047*	.33

*$p \leq .05$.
**$p \leq .01$.
***$p \leq .001$ (Bonferroni Correction threshold).

Table 7. Significant results of Pearson's Chi square test of categorical programmatic and educator variables for top and bottom quartile programs.

Characteristic	X²	p	Φ
Identity: Friend	7.39	0.007**	0.223
Identity: Walking encyclopedia	11.12	< 0.001***	−0.273
Pace good	9.84	0.002**	0.257
Pace slow	11.55	< 0.001***	−0.278

*$p \leq .05$.
**$p \leq .01$.
***$p \leq .001$ (Bonferroni Correction threshold).

Table 8. Significant results of independent samples t-test comparing educator characteristic observational scores between top and bottom quartiles.

Educator characteristic	Top quartile m (sd)	Bottom quartile M (SD)	t	df	p	Cohen's d
Emotional support index	3.19 (.48)	2.82 (.58)	4.276	147	<.001***	.71
Responsiveness	3.29 (.61)	2.84 (.70)	4.226	147	<.001***	.70
Comfort and clarity index	3.42 (.42)	3.15 (.48)	3.586	144	<.001***	.60
Formality	2.59 (.57)	2.81 (.54)	−2.458	147	0.015*	.41
Audibility	3.55 (.60)	3.31 (.62)	2.365	147	0.019*	.39

*$p \leq .05$.
**$p \leq .01$.
***$p \leq .001$ (Bonferroni Correction threshold).

($p \leq 0.001$), we conducted a special type of multiple regression known as relative weight analysis (RWA) (Johnson 2000) using RWA-Web (Tonidandel and LeBreton 2015). Because RWA is a type of regression that tests linear relationships with a continuous dependent variable, we selected the eleven variables with the strongest bivariate relationships with GMC EE21 (rather than the top vs. bottom quartile comparisons). RWA also examines the relative importance of each predictor within the regression. The results indicated that a weighted linear combination of the 11 most predictive programmatic, educator, and setting variables explained 18% of the variance in EE21 (Table 10.). An examination of the relative weights revealed that *group size, educator comfort and clarity, responsiveness, naturalness of setting, and novelty of setting* explained a statistically unique and significant amount of variance in GMC EE21 (Table 10.). RWA identifies the incremental contribution of each variable in the presence of other correlated predictors (Tonidandel and LeBreton 2011). The relative weights (RW) suggest that *group size* (.044), *naturalness of the setting* (.024), *novelty of the setting* (.022), responsiveness (.0154), and *educator comfort and clarity* (.153) appear to explain unique incremental variance

Table 9. Significant results of independent samples t-test comparing setting characteristic observational scores between top and bottom quartiles.

Setting characteristic	Top quartile M (SD)	Bottom quartile M (SD)	t	df	p	Cohen's D
Naturalness	2.85 (.73)	2.32 (.80)	−4.231	147	<.001***	.70
Novelty	2.76 (.82)	2.30 (.66)	−3.801	147	<.001***	.63
Beauty	2.72 (.65)	2.45 (.66)	−2.356	138	0.020*	.40
Immersion	2.43 (.79)	2.14 (.75)	−2.313	147	0.022*	.38

*$p \leq .05$.
**$p \leq .01$.
***$p \leq .001$ (Bonferroni Correction threshold).

Table 10. Results of relative weight analysis.

Predictor	RW	CI-L	CI-U	RS-RW (%)
Group Size[a]	.0437	.0101	.0992	23.77
Naturalness[a]	.0239	.0024	.0718	13.01
Novelty[a]	.0224	.001	.0692	12.21
Responsive[a]	.0154	.0004	.0484	8.41
Educator Comfort and Clarity Index[a]	.0153	.0005	.0485	8.35
Staging	.0161	−.002	.065	8.74
Emotional Support Index	.0128	−.0009	.0419	6.99
Verbal engagement	.0123	−.0029	.0495	6.70
Place-based	.0107	−.0029	.0437	5.83
Question quality	.0095	−.0028	.0421	5.18
Transitions	.0015	−.0136	.0164	0.81

RW = raw relative weight (sum of RW = R2); CI-L = lower bound of confidence interval used to test the statistical significance of RW; CI-U = upper bound of confidence interval used to test the statistical significance of RW; RS-RW = relative weight rescaled as a percentage of predicted variance in the criterion variable attributed to each predictor (Tonidandel and LeBreton 2015).
[a]Significant at 95% (CI did not contain 0).

in GMC EE21, while the other six predictor variables do not explain unique variance because of their correlations with the other predictors (Table 10) (Tonidandel and LeBreton 2011, 2015).

Discussion

Our study attempted to identify the techniques and approaches that most consistently lead to better outcomes in 299 diverse 5th–8th grade field trip programs across the United States. We first examined the relationship between different programmatic, educator, and setting characteristics with student outcomes. Second, we examined which of these characteristics best distinguished between the programs yielding the most positive and least positive student outcomes. Finally, we conducted RWA to determine the proportion of variance in the overall GMC EE21 outcome measure that could be explained by the characteristics with strongest predictive ability in our sample. We highlight key lessons that emerged from our observations and these statistical explorations below.

Our initial descriptive results suggest that EE field trip programs largely appear to be taught by educators with high levels of confidence in performing their duties and high degrees of emotional support for their students. The results also indicate that many of the most commonly promoted techniques in EE (see NAAEE 2017, 2020) were rarely encountered in our sample of 299 single-day field trip programs, including individual and group reflection, incorporating multiple viewpoints, group discussion, and teaching 21st century skills. One potential explanation for this is that these field trip programs were focused on meeting educational standards, which can come at the expense of using some of the techniques and pedagogies promoted by the field. Some argue, including Gruenewald and Manteaw (2007) and Stevenson (2007), that the influence of 'No Child Left Behind Act' and the focus on meeting educational standards has eroded the practice of EE in the U.S. and its mission to enhance environmental literacy and

address key environmental issues. Research on teacher motivations for attending field trip programs supports this argument; most teachers desire to link activities to curriculum standards (e.g. Kisiel 2005; Storksdieck 2006). Our results also suggest that many EE field trips in the U.S. are strongly influenced by formal science education practices and some observed programs performed activities that were roughly equivalent to classroom science lab assignments, rather than providing fully immersive, place-based experiences espoused by EE experts and practitioners (Krasny 2020; Stern, Powell, and Hill 2014; Sobel 2008; NAAEE 2020). Our results suggest that these programs can still achieve positive learning outcomes. However, certain practices, as expected, were linked with better outcomes than others.

Which programmatic, educator, and setting characteristics are associated with better outcomes?

The series of complementary tests we conducted identified 11 programmatic, educator, and setting characteristics that were most powerfully and consistently associated with positive learning outcomes (Table 5). These characteristics were also highly related to each other. For example, educators that provide high degrees of responsiveness also tended to provide high levels of emotional support (Table 5). The RWA revealed that five of these 11 variables accounted for unique incremental variance; the other six did not account for additional unique variance and were largely accounted for by these five. Overall, the RWA suggests that these variables explain 18% of the variance in learning outcomes, while controlling for grade and race (GMC EE21).

In this study, smaller groups exhibited more positive outcomes. Prior research has commonly found that smaller class sizes are associated with increased academic achievement in formal education (Shin and Chung 2009; Bosworth 2014; Chingos and Whitehurst 2011). However, research on group size has been limited in the case of EE field trips (Bitgood 1989; Stern, Powell, and Ardoin 2008) and similar experiences (Powell, Kellert, and Ham 2009; Stern and Powell 2013) and findings have been less consistent. We know of no study that has examined the influence of group size on single-day EE field trips on student outcomes. Our study suggests that breaking larger groups into smaller, more intimate groups are likely beneficial for single-day EE field trip programs. Further explorations of our data did not yield an optimal group size. Rather, we found a roughly linear relationship.

Our results also suggest programs that occurred in more natural and more novel natural settings were associated with more positive learning outcomes. Natural settings can range from urban gardens to wilderness areas (e.g. Dale et al. 2020). Novel natural settings may include unique natural phenomena or viewing and interacting with wildlife, particularly unique and charismatic fauna not commonly observed by the program participants otherwise (e.g. Skibins, Powell, and Hallo 2013). These results mirror prior research in formal (e.g. Kuo, Barnes, and Jordan 2019) and informal education (e.g. Browning and Rigolon 2019) and support widely held assumptions in the field of the importance of connecting participants with nature. However, these results do not suggest that programs in more urban locations cannot be effective. Rather, whether in an urban or rural environment, educators should consider how to incorporate natural settings and novelty in their programs to the greatest extent possible.

Related to these findings, our results also suggest that programs that focused on the unique natural, cultural, and social attributes of the location in their programming were associated with more positive learning outcomes. Highlighting the unique components of a location is a hallmark of place-based education and highly promoted in EE (NAAEE 2020; Krasny 2020; Stern, Powell, and Hill 2014; Sobel 2008; Gruenewald 2003, 2008). Many programs in our sample largely ignored the unique attributes of the place and instead provided content and activities that could occur in any location. These programs that did not or minimally highlighted the attributes of the location, on average, yielded less positive student outcomes.

Our results also highlighted the important role that the educator plays in delivering a successful EE program. Specifically, our study found that educators that (1) clearly communicated both

instructions and content to their students, (2) exuded comfort, confidence, and a good working knowledge of the location and content, and (3) created an emotionally supportive and responsive class environment produced more positive student outcomes. Research in both formal education (e.g. Stronge, Ward, and Grant 2007, 2011) as well as more informal settings such as interpretation in parks have found similar results (Stern and Powell 2013). In particular, acknowledging student success; affinity seeking behaviors, such as providing eye-contact, smiling, and listening; providing positive feedback; and responding to students' questions, interests, and non-verbal cues all build a positive learning environment and trust between students and an educator (e.g. Pianta, La Paro, and Hamre 2008, Pianta and Hamre; Rudasill, Gallagher, and White). Formal education research has extensively explored the importance of these relational and positive learning environment skills for teachers (e.g. Pianta and Hamre 2009; Rudasill, Gallagher, and White 2010; Reyes et al. 2012) and our work further reinforces this importance in more non-formal educational settings.

The results of this study also suggest the importance of using quality transitions to connect different concepts, programmatic elements/activities, and locations, as well as providing a conclusion at the end of the program that reinforces the take-home message. In the context of EE field trip programs, transitions can be used to maintain cognitive engagement by connecting content and concepts, by linking locations (e.g. 'students, I want you to pay attention to how the landscape changes as we walk to our next location. I'm going to ask you about what you notice when we arrive'), or by connecting different activities to help students to draw larger lessons. Providing a well-crafted conclusion at the end of a field trip program serves as a way to summarize the key themes and the learning that occurred, as well as offer opportunities for students to reflect upon the experience (Ham 2013, 1992). Research in interpretation (Ham 2013; Powell and Stern 2013; Stern and Powell 2013) reinforces these results and speaks to the need to plan, organize, and deliver day-long EE field trip programs as a cohesive, well-organized, and integrated whole.

Our research suggests that verbal engagement and high-quality questions, in particular, also enhance student outcomes. Prior research and theory outline what constitutes a 'high quality' question: it is one that cannot be answered with yes or no; challenges students to apply new information: explores relationships and cause and effect; draw comparisons; elicits opinions and evaluation; focuses attention; and/or drives inference or problem solving (Ham 1992; NPS 2019). Our results suggest that any verbal engagement can enhance program outcomes for students, but high-quality questions during an EE field trip program raises the potential to increase cognitive engagement, critical thinking, and meta-cognition in students.

The state in which students arrive at a field trip program (staging) also influences learning outcomes. Did the students arrive disorganized, late, or confused or were they well-prepared and ready to participate in the field trip? While being late can often be outside the control of visiting school groups or program providers, other forms of preparation can be addressed. Extensive research on field trip programs, including EE field trip programs, suggests that student preparation, including logistical preparation, before a program is important for setting the stage for effective student learning (e.g. Storksdieck 2006; DeWitt and Storksdieck 2008; Lee, Stern, and Powell 2020). In some cases, students might be exposed to a pre-visit video about the site, which can focus on the right gear to bring, clarify expectations for the visit, introduce important subject matter, vocabulary, tools, or techniques that might help to prepare them for a more fruitful exhibit. Alternatively, teachers might cover these topics in the classroom, sometimes with the help of support materials designed by the field trip providers. Prior research suggests that teachers sometimes forgo preparing for a field trip due to a lack of familiarity with the site and/or a perceived lack of time in the classroom (e.g. Anderson, Kisiel, and Storksdieck 2006; Cox-Petersen and Pfaffinger 1998). Developing stronger partnerships with visiting school teachers could potentially enhance preparation for both teachers and students (Lee, Stern and Powell 2020) and enhance staging.

Finally, several other programmatic and educator characteristics were also associated with higher performing programs and had mid-level effect sizes (above .5) (Cohen 1992). Effective class management was positively associated with outcomes. Meanwhile, three characteristics

were negatively associated with outcomes: too slow of program pace, providing lecture-based programming, and educators that operated as walking encyclopedias regurgitating facts. Prior research in formal education supports these findings that effective class management improves academic performance (e.g. Pianta and Hamre 2009). Research in informal settings also supports the findings that an overreliance on didactic lecture based programming and fact-based approaches erode outcomes (e.g. Stern and Powell 2013).

Limitations

Our study and analyses identified program characteristics that were most consistently linked to short-term student outcomes across a large and diverse sample of single-day EE field trip programs. The final RWA model accounted for 18% of the variance in the GMC EE21 outcome. Thus, although these 11 characteristics were consistently related to better student outcomes, approximately 80% of the variance was unexplained and likely attributable to other factors, such as pre-visit preparation (Lee, Stern and Powell 2020), students' backgrounds (Stern, Powell and Frensley 2022), measurement shortcomings, and other factors not measured in this study. Moreover, immediate post program measurement of student outcomes provided a consistent means for comparison across programs, however the study was unable to measure longer-term outcomes for participating students. Thus, the findings distinguish the approaches and characteristics that influence immediate outcomes, but we cannot necessarily extrapolate to what may influence longer-term impacts. While conducting a large-scale study of this nature using pre-experience, immediate post experience, and follow-up outcome measures would be a challenging logistical task, it could yield more definitive results regarding the program characteristics that lead to more lasting impacts. Moreover, these studies could better examine the influence of pre-experience preparation and follow-up activities (Lee, Stern and Powell 2020; Smith-Sebasto and Cavern 2006).

Because we so rarely observed some program characteristics, such as reflection, multiple viewpoints, storytelling, and others, our study was unable to assess their potential relationships with student outcomes. It may be that these approaches are particularly rare in the types of EE we observed—that is, single-day field trips for middle school students. Observing other types of EE programs, such as multi-day residential programs, might yield greater insights into these and other characteristics. We encourage future researchers to build off our work and the work of others to examine similar programs in other contexts, as well as virtual programs and multi-day residential programs in experimental or comparative ways to begin to better understand what works in these other EE contexts (for example, see Merritt et al. 2022; Frensley et al. 2022).

Because we aggregated the data to the program level, our study also did not isolate which program approaches are best for different audiences (specific grade/age, races and ethnicities, or sexual identity), but rather which practices showed strongest relationships with outcomes across all audiences. Because prior research suggests that EE is not one-size fits all, future research should examine what works best for different audiences and in what contexts. Aggregating the data to the program level precluded us from using some of the individual-student demographic data. Our aggregated sample size was also insufficient to examine which characteristics tended to function better or worse within meaningful aggregated subsamples. Using multi-level modeling may allow for identifying the relationship between different approaches and the outcomes for specific types of individuals. However, a larger and more diverse sample size would be necessary to enable such cross-context investigations. Moreover, our quantitative measures, while broadly descriptive, do not address the nuances that a more qualitative examination of fewer programs would be able to explore. We urge future researchers to pursue comparative approaches that combine qualitative observations of programs with quantitative outcome measures to provide holistic descriptions of extraordinary individual programs.

Conclusion

Our study observed a large sample of single-day field trip programs for youth in grades 5th–8th in the US in an attempt to identify the techniques and approaches that were most consistently associated with more positive learning outcomes. The results do not provide an exclusive list of approaches and characteristics that definitively enhance learning outcomes—as our study variables were limited to those we observed within our sample. However, as one of the first studies of this kind, we urge other researchers to build upon our research design to further ground the field of EE in evidence and better adapt to the changing demographics and lived experiences of youth worldwide.

ORCID

Robert B. Powell http://orcid.org/0000-0003-2775-2571
Marc J. Stern http://orcid.org/0000-0002-0294-8941
Brandon Troy Frensley http://orcid.org/0000-0002-4670-5503

References

Anderson, D., J. Kisiel, and M. Storksdieck. 2006. "Understanding Teachers' Perspectives on Field Trips: Discovering Common Ground in Three Countries." *Curator: The Museum Journal* 49 (3): 365–386. doi:10.1111/j.2151-6952.2006.tb00229.x

Archer, D., and S. Wearing. 2003. "Self, Space, and Interpretive Experience: The Interactionism of Environmental Interpretation." *Journal of Interpretation Research* 8 (1): 7–23. doi:10.1177/109258720300800102

Ardoin, N. M. 2006. "Toward an Interdisciplinary Understanding of Place: Lessons for Environmental Education." *Canadian Journal of Environmental Education (CJEE)* 11 (1): 112–126.

Ardoin, N. M., K. Biedenweg, and K. O'Connor. 2015. "Evaluation in Residential Environmental Education: An Applied Literature Review of Intermediary Outcomes." *Applied Environmental Education & Communication* 14 (1): 43–56. doi:10.1080/1533015X.2015.1013225

Ardoin, N. M., A. W. Bowers, N. W. Roth, and N. Holthuis. 2018. "Environmental Education and K-12 Student Outcomes: A Review and Analysis of Research." *The Journal of Environmental Education* 49 (1): 1–17. doi:10.1080/00958964.2017.1366155

Bamberger, Y., and T. Tal. 2007. "Learning in a Personal Context: Levels of Choice in a Free Choice Learning Environment in Science and Natural History Museums." *Science Education* 91 (1): 75–95. doi:10.1002/sce.20174

Beck, L., and T. T. Cable. 2002. *Interpretation for the 21st Century: Fifteen Guiding Principles for Interpreting Nature and Culture* (2nd ed.). Champaign: Sagamore.

Bitgood, S. 1989. "School Field Trips: An Overview." *Visitor Behavior* 4 (2): 3–6.

Browning, M. H., and D. H. Locke. 2020. "The Greenspace-Academic Performance Link Varies by Remote Sensing Measure and Urbanicity around Maryland Public Schools." *Landscape and Urban Planning* 195: 103706. doi:10.1016/j.landurbplan.2019.103706

Browning, M. H., and A. Rigolon. 2019. "School Green Space and Its Impact on Academic Performance: A Systematic Literature Review." *International Journal of Environmental Research and Public Health* 16 (3): 429. doi:10.3390/ijerph16030429

Born, R. J. G. van den, R. H. J. Lenders, W. T. de Groot, and E. Huijsman. 2001. "The New Biophilia: An Exploration of Visions of Nature in Western Countries." *Environmental Conservation* 28 (1): 65–75. doi:10.1017/S0376892901000066

Bosworth, R. 2014. "Class Size, Class Composition, and the Distribution of Student Achievement." *Education Economics* 22 (2): 141–165. doi:10.1080/09645292.2011.568698

Bourke, N., C. Buskist, and J. Herron. 2014. "Residential Environmental Education Center Program Evaluation: An Ongoing Challenge." *Applied Environmental Education & Communication* 13 (2): 83–90. doi:10.1080/1533015X.2014.944632

Bowers, E. P., Y. Li, M. K. Kiely, A. Brittian, J. V. Lerner, and R. M. Lerner. 2010. "The Five Cs Model of Positive Youth Development: A Longitudinal Analysis of Confirmatory Factor Structure and Measurement Invariance." *Journal of Youth and Adolescence* 39 (7): 720–735. doi:10.1007/s10964-010-9530-9

Budescu, D. V. 1993. "Dominance Analysis: A New Approach to the Problem of Relative Importance of Predictors in Multiple Regression." *Psychological Bulletin* 114 (3): 542–551. doi:10.1037/0033-2909.114.3.542

Budescu, D. V., and R. Azen. 2004. "Beyond Global Measures of Relative Importance: Some Insights from Dominance Analysis." *Organizational Research Methods* 7 (3): 341–350. doi:10.1177/1094428104267049

Chandrasekhar, S. 1987. *Truth and Beauty: Aesthetics and Motivations in Science*. Chicago, IL: University of Chicago Press. doi:10.1016/j.palwor.2007.07.002.

Cohen, J. 1992. "Statistical Power Analysis." *Current Directions in Psychological Science* 1 (3): 98–101. doi:10.1111/1467-8721.ep10768783

Chingos, M. M., and G. J. Whitehurst. 2011. *Class Size: What Research Says and What It Means for State Policy*. Washington, DC: Brookings Institute.

Cox-Petersen, Anne M., and Julie A. Pfaffinger. 1998. "Teacher Preparation and Teacher-Student Interactions at a Discovery Center of Natural History." *Journal of Elementary Science Education* 10 (2): 20–35. doi:10.1007/BF03173782

Dale, R. G., R. B. Powell, M. J. Stern, and B. A. Garst. 2020. "Influence of the Natural Setting on Environmental Education Student Outcomes." *Environmental Education Research* 26 (5): 613–631. doi:10.1080/13504622.2020.1738346

de Waal, F. B. M. 2008. "Putting the Altruism Back into Altruism: The Evolution of Empathy." *Annual Review of Psychology* 59 (1): 279–300. doi:10.1146/annurev.psych.59.103006.093625

DeVellis, R. F. 2003. *Scale Development: Theory and Applications Applied Social Research Applications*. 2nd ed. Thousand Oaks, CA: Sage Publishing.

DeWitt, J., and M. Storksdieck. 2008. "A Short Review of School Field Trips: Key Findings from the past and Implications for the Future." *Visitor Studies* 11 (2): 181–197. doi:10.1080/10645570802355562

Eisenberg, N., R. Shell, J. Pasternack, R. Lennon, R. Beller, and R. M. Mathy. 1987. "Prosocial Development in Middle Childhood: A Longitudinal Study." *Developmental Psychology* 23 (5): 712–718. doi:10.1037/0012-1649.23.5.712

Fenichel, M., and H. A. Schweingruber. 2010. "Surrounded by Science: Learning Science in Informal Environments." *Board on Science Education, Center for Education, Division of Behavioral and Social Sciences and Education*. Washington, DC: The National Academies Press.

Frensley, B. T., M. J. Stern, and R. B. Powell. 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. doi:10.1080/00958964.2020.1727404

Frensley, B. T., M. J. Stern, R. B. Powell, and M. G. Sorice. 2022. "Investigating the Links between Student Engagement, Self-Determination, and Environmental Literacy at a Residential Environmental Education Center." *Environmental Education Research*. 53 (4): 186–198. doi:10.1080/13504622.2022.2044454

Frauman, E., and W. C. Norman. 2003. "Managing Visitors via 'Mindful' Information Services: One Approach in Addressing Sustainability." *Journal of Park and Recreation Administration* 21 (4): 87–104.

Gage, N. L., and M. C. Needels. 1989. "Process-Product Research on Teaching: A Review of Criticisms." *The Elementary School Journal* 89 (3): 253–300. doi:10.1086/461577

Garst, B. A. 2018. "Nature and Youth Development." In *Youth Development: Principles and Practices in out of School Settings*, edited by P. A. Witt and L. L. Caldwell, 241–268. Urbana, IL: Sagamore-Venture.

Garst, B. A., L. P. Browne, and M. D. Bialeschki. 2011. "Youth Development and the Camp Experience." *New Directions for Youth Development* 2011 (130): 73–87. doi:10.1002/yd.398

Gruenewald, D. A. 2003. "The Best of Both Worlds: A Critical Pedagogy of Place." *Educational Researcher* 32 (4): 3–12. doi:10.3102/0013189X032004003

Gruenewald, D. A. 2008. "The Best of Both Worlds: A Critical Pedagogy of Place." *Environmental Education Research* 14 (3): 308–324. doi:10.1080/13504620802193572

Gruenewald, D. A., and B. O. Manteaw. 2007. "Oil and Water Still: How No Child Left behind Limits and Distorts Environmental Education in US Schools." *Environmental Education Research* 13 (2): 171–188. doi:10.1080/13504620701284944

Ham, S. H. 1992. *Environmental Interpretation: A Practical Guide for People with Big Ideas and Small Budgets*. Golden, CO: Fulcrum Publishing.

Ham, S. H. 2009. "From Interpretation to Protection: Is There a Theoretical Basis?" *Journal of Interpretation Research* 14 (2): 49–57. doi:10.1177/109258720901400204

Ham, S. H. 2013. *Interpretation: Making a Difference on Purpose*. Golden, CO: Fulcrum Publishing.

Ham, S. H., and B. M. Weiler. 2002. "Tour Guide Training: A Model for Sustainable Capacity Building in Developing Countries." *Journal of Sustainable Tourism* 10 (1): 52–69.

Hamre, B. K., and R. C. Pianta. 2005. "Can Instructional and Emotional Support in the First-Grade Classroom Make a Difference for Children at Risk of School Failure?" *Child Development* 76 (5): 949–967. doi:10.1111/j.1467-8624.2005.00889.x

Holton, G. J. 1988. *Thematic Origins of Scientific Thought: Kepler to Einstein*. Cambridge, MA: Harvard University Press

Institute of Museum and Library Services. 2009. *Museums, Libraries, and 21st Century Skills*. Washington, DC: Library of Congress. doi:10.1037/e483242006-005

Jacobson, S. K. 1999. *Communication Skills for Conservation Professionals*. Washington, DC: Island Press.

Jarvis, C. B., MacKenzie, S. B., and Podsakoff, P. M. 2003. "A Critical Review of Construct Indicators and Measurement Model Misspecification in Marketing and Consumer Research." *Journal of Consumer Research* 30 (2): 199–218.

Johnson, J. W. 2000. "A Heuristic Method for Estimating the Relative Weight of Predictor Variables in Multiple Regression." *Multivariate Behavioral Research* 35 (1): 1–19. doi:10.1207/S15327906MBR3501_1

Johnson, J. W. 2004. "Factors Affecting Relative Weights: The Influence of Sampling and Measurement Error." *Organizational Research Methods* 7 (3): 283–299. doi:10.1177/1094428104266018

Jose, S., P. G. Patrick, and C. Moseley. 2017. "Experiential Learning Theory: The Importance of Outdoor Classrooms in Environmental Education." *International Journal of Science Education, Part B* 7 (3): 269–284. doi:10.1080/21548455.2016.1272144

Kahn, P. H., Jr., and Kellert, S. R., eds. 2002. *Children and Nature: Psychological, Sociocultural, and Evolutionary Investigations*. Cambridge, MA: MIT Press.

Kaplan, R., S. Kaplan, and R. Ryan. 1998. *With People in Mind: Design and Management of Everyday Nature*. Washington, DC: Island Press.

Kellert, S. R. 2002. "Experiencing Nature: Affective, Cognitive, and Evaluative Development in Children." In *Children and Nature. Psychological, Sociocultural, and Evolutionary Investigations*, edited by P. H. Kahn & S. R. Kellert, 117–151. Cambridge, MA: MIT Press.

Kellert, S. R. 2005. "Building for Life." *Designing and Understanding the Human–Nature Connection*. Washington, DC: Island Press.

Kellert, S. R. 2008. "A Biocultural Basis for an Ethic Toward the Natural Environment." In *Foundations of Environmental Sustainability: The Coevolution of Science and Policy*, edited by L. Rockwood, R. Stewart, and T. Dietz, 321–332. Oxford, UK: Oxford University Press.

Keltner, D., A. Kogan, P. K. Piff, and S. R. Saturn. 2014. "The Sociocultural Appraisals, Values, and Emotions (save) Framework of Prosociality: Core Processes from Gene to Meme." *Annual Review of Psychology* 65 (1): 425–460. doi:10.1146/annurev-psych-010213-115054

Kisiel, J. 2005. "Understanding Elementary Teacher Motivations for Science Fieldtrips." *Science Education* 89 (6): 936–955. doi:10.1002/sce.20085

Kline, R. B. 2005. *Principles and Practice of Structural Equation Modeling*. 2nd ed. New York: The Guilford Press.

Knapp, D., and L. Yang. 2002. "A Phenomenological Analysis of Long-Term Recollections of an Interpretive Program." *Journal of Interpretation Research* 7 (2): 7–17. doi:10.1177/109258720200700202

Knudson, D. M., T. T. Cable, and L. Beck. 2003. *Interpretation of Cultural and Natural Resources*. 2nd ed. State College: Venture Publishing.

Kohlberg, L. 1971. "Stages of Moral Development." *Moral Education* 1 (51): 23–92.

Krasny, M. E. 2020. *Advancing Environmental Education Practice*. Ithaca, NY: Cornell University Press.

Kuo, M., M. Barnes, and C. Jordan. 2019. "Do Experiences with Nature Promote Learning? Converging Evidence of a Cause-and-Effect Relationship." *Frontiers in Psychology* 10: 305. doi:10.3389/fpsyg.2019.00305

Larsen, D. L. 2003. *Meaningful Interpretation: How to Connect Hearts and Minds to Places, Objects, and Other Resources*. Fort Washington, PA: Eastern National.

Lee, H., M. J. Stern, and R. B. Powell. 2020. "Assessing the Influence of Preparation and Follow-up on Student Outcomes Associated with Environmental Education Field Trips." *Environmental Education Research* 26 (7): 989–1007. doi:10.1080/13504622.2020.1765991

Lewis, W. J. 2005. *Interpreting for Park Visitors*. 9th ed. Fort Washington, PA: Eastern National.

Lerner, Richard M., Jacqueline V. Lerner, Jason B. Almerigi, Christina Theokas, Erin Phelps, Steinunn Gestsdottir, Sophie Naudeau, et al. 2005. "Positive Youth Development, Participation in Community Youth Development Programs, and Community Contributions of Fifth-Grade Adolescents: Findings from the First Wave of the 4-H Study of Positive Youth Development." *The Journal of Early Adolescence* 25 (1): 17–71. doi:10.1177/0272431604272461

Merritt, E. G., M. J. Stern, R. B. Powell, and B. T. Frensley. 2022. "A Systematic Literature Review to Identify Evidence-Based Principles to Improve Online Environmental Education." *Environmental Education Research* 28 (5): 674–694. doi:10.1080/13504622.2022.2032610

Merritt, E. G., S. B. Wanless, S. E. Rimm-Kaufman, C. Cameron, and J. L. Peugh. 2012. "The Contribution of Teachers' Emotional Support to Children's Social Behaviors and Self-Regulatory Skills in First Grade." *School Psychology Review* 41 (2): 141–159. doi:10.1080/02796015.2012.12087517

Moscardo, G. 1999. *Making Visitors Mindful: Principles for Creating Quality Sustainable Visitor Experiences through Effective Communication*. Champaign, IL: Sagamore Publishing.

Moseley, C., H. Summerford, M. Paschke, C. Parks, and J. Utley. 2020. "Road to Collaboration: Experiential Learning Theory as a Framework for Environmental Education Program Development." *Applied Environmental Education & Communication* 19 (3): 238–258. doi:10.1080/1533015X.2019.1582375

North American Association for Environmental Education (NAAEE). 2017. *Community Engagement: Guidelines for Excellence*. Washington, DC: North American Association for Environmental Education.

North American Associate of Environmental Education (NAAEE). 2020. *Guidelines for Excellence Series*. Washington, DC: North American Association for Environmental Education.

Nathans, L. L., F. L. Oswald, and K. Nimon. 2012. "Interpreting Multiple Linear Regression: A Guidebook of Variable Importance." *Practical Assessment, Research & Evaluation* 17 (9): 1–19.

National Park Service. 2014. *Achieving Relevance in Our Second Century*. Washington, DC: National Park Service.

National Park Service. 2019. *Forging Connections through Audience Centered Experiences*. Harpers Ferry, WV: Stephen T. Mather Training Center, Interpretive Development Program.

National Research Council (NRC), Board on Science Education, Center for Education, Division of Behavioral and Social Sciences and Education. 2009. "Learning Science in Informal Environments: People, Places, and Pursuits." In *Committee on Learning Science in Informal Environments*, edited by P. Bell, B. Lewenstein, A. W. Shouse, and M. A. Feder. Washington, DC: The National Academies Press.

O'Hare, A., R. B. Powell, M. J. Stern, and E. P. Bowers. 2020. "Influence of Educator's Emotional Support Behaviors on Environmental Education Student Outcomes." *Environmental Education Research* 26 (11): 1556–1577. doi:10.1080/13504622.2020.1800593

Pianta, R. C., and B. K. Hamre. 2009. "Conceptualization, Measurement, and Improvement of Classroom Processes: Standardized Observation Can Leverage Capacity." *Educational Researcher* 38 (2): 109–119. doi:10.3102/0013189X09332374

Pianta, R. C., K. M. La Paro, and B. K. Hamre. 2008. *Classroom Assessment Scoring System: Manual K-3*. Baltimore, MD: Paul H Brookes Publishing.

Piaget, J. 1936. *Origins of Intelligence in the Child*. London: Routledge & Kegan Paul.

Powell, R. B., S. R. Kellert, and S. H. Ham. 2009. "Interactional Theory and the Sustainable Nature-Based Tourism Experience." *Society & Natural Resources* 22 (8): 761–776. doi:10.1080/08941920802017560

Powell, R. B., and M. J. Stern. 2013. "Is It the Program or the Interpreter? Modeling the Influence of Program Characteristics and Interpreter Attributes on Visitor Outcomes." *Journal of Interpretation Research* 18 (2): 45–60. doi:10.1177/109258721301800203

Powell, R. B., M. J. Stern, B. T. Frensley, and D. Moore. 2019. "Identifying and Developing Crosscutting Environmental Education Outcomes for Adolescents in the 21st Century (EE21)." *Environmental Education Research* 25 (9): 1281–1299. doi:10.1080/13504622.2019.1607259

Podsakoff, P. M., S. B. MacKenzie, J. Y. Lee, and N. P. Podsakoff. 2003. "Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies." *Journal of Applied Psychology* 88 (5): 879–903. doi:10.1037/0021-9010.88.5.879

Ragin, C. C. 2008. *Redesigning Social Inquiry: Fuzzy Sets and beyond*. Chicago: University of Chicago Press.

Reyes, M. R., M. A. Brackett, S. E. Rivers, M. White, and P. Salovey. 2012. "Classroom Emotional Climate, Student Engagement, and Academic Achievement." *Journal of Educational Psychology* 104 (3): 700–712. doi:10.1037/a0027268

Rickinson, M. 2001. "Learners and Learning in Environmental Education: A Critical Review of the Evidence." *Environmental Education Research* 7 (3): 207–320. doi:10.1080/13504620120065230

Rudasill, K. M., K. C. Gallagher, and J. M. White. 2010. "Temperamental Attention and Activity, Classroom Emotional Support, and Academic Achievement in Third Grade." *Journal of School Psychology* 48 (2): 113–134. doi:10.1016/j.jsp.2009.11.002

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

Ryan, R. M., N. Weinstein, J. Bernstein, K. W. Brown, L. Mistretta, and M. Gagne. 2010. "Vitalizing Effects of Being Outdoors and in Nature." *Journal of Environmental Psychology* 30 (2): 159–168. doi:10.1016/j.jenvp.2009.10.009

Shin, I. S., and J. Y. Chung. 2009. "Class Size and Student Achievement in the United States: A Meta-Analysis." *KEDI Journal of Educational Policy* 6 (2): 3–19.

Skibins, J. C., R. B. Powell, and J. C. Hallo. 2013. "Charisma and Conservation: Charismatic Megafauna's Influence on Safari and Zoo Tourists' Pro-Conservation Behaviors." *Biodiversity and Conservation* 22 (4): 959–982. doi:10.1007/s10531-013-0462-z

Smith-Sebasto, N. J., and L. Cavern. 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. doi:10.3200/JOEE.37.4.3-17

Sobel, D. 2008. *Childhood and Nature: Design Principles for Educators*. Portland, ME: Stenhouse Publishers.

Stern, M. J., and R. B. Powell. 2013. "What Leads to Better Visitor Outcomes in Live Interpretation?" *Journal of Interpretation Research* 18 (2): 9–43. doi:10.1177/109258721301800202

Stern, M. J., and R. B. Powell. 2020. "Field Trips and the Experiential Learning Cycle." *Journal of Interpretation Research* 25 (1): 46–50. doi:10.1177/1092587220963530

Stern, M. J., R. B. Powell, and N. M. Ardoin. 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. doi:10.3200/JOEE.39.4.31-43

Stern, M. J., R. B. Powell, and D. Hill. 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. doi:10.1080/13504622.2013.838749

Stern, M. J., R. B. Powell, and B. T. Frensley. 2022. "How Do Environmental Education Field Trips for Adolescent Youth in the United States Influence Audiences of Different Grade Level, Race, and Socioeconomic Class?" *Environmental Education Research* 28 (2): 197–215. doi:10.1080/13504622.2021.1990865

Stevenson, R. B. 2007. "Schooling and Environmental Education: Contradictions in Purpose and Practice." *Environmental Education Research* 13 (2): 139–153. doi:10.1080/13504620701295726

Stronge, J. H., T. J. Ward, and L. W. Grant. 2011. "What Makes Good Teachers Good? A Cross-Case Analysis of the Connection between Teacher Effectiveness and Student Achievement." *Journal of Teacher Education* 62 (4): 339–355. doi:10.1177/0022487111404241

Stronge, J. H., T. J. Ward, P. D. Tucker, and J. L. Hindman. 2007. "What is the Relationship between Teacher Quality and Student Achievement? An Exploratory Study." *Journal of Personnel Evaluation in Education* 20 (3–4): 165–184. doi:10.1007/s11092-008-9053-z

Storksdieck, M. 2006. *Field Trips in Environmental Education*. Munich, Germany: BWV Verlag.

Tabachnick, B. G., and L. S. Fidell. 2018. *Using Multivariate Statistics*. 7th ed. Boston, MA: Pearson Education, Inc.

Thompson, J. L., and A. K. Houseal. 2020. *America's Largest Classroom: What We Learn from Our National Parks*. Oakland, CA: University of California Press.

Tilden, F. 1957. *Interpreting Our Heritage*. Chapel Hill, NC: University of North Carolina Press.

Tonidandel, S., J. M. LeBreton, and J. W. Johnson. 2009. "Determining the Statistical Significance of Relative Weights." *Psychological Methods* 14 (4): 387–399.

Tonidandel, S., and J. M. LeBreton. 2011. "Relative Importance Analysis: A Useful Supplement to Regression Analysis." *Journal of Business and Psychology* 26 (1): 1–9. doi:10.1007/s10869-010-9204-3

Tonidandel, S., and J. M. LeBreton. 2015. "RWA Web: A Free, Comprehensive, Web-Based, and User-Friendly Tool for Relative Weight Analyses." *Journal of Business and Psychology* 30 (2): 207–216. doi:10.1007/s10869-014-9351-z

UNESCO. 1977. *The Tbilisi Declaration. Intergovernmental Conference on Environmental Education*, 14–26. Tbilisi, Georgia: UNESCO.

USDA (United States Department of Agriculture) Food and Nutrition Services. 2020. https://www.fns.usda.gov/.

Wallace, G. N., and Gaudry, C. J. 2002. "An Evaluation of the 'Authority of the Resource' Interpretive Technique by Rangers in Eight Wilderness/Backcountry Areas." *Journal of Interpretation Research*, 7: 43–68.

Ward, C. W., and Wilkinson, A. E. 2006. *Conducting Meaningful Interpretation: A Field Guide for Success*. Golden: Fulcrum.

Wells, N. M. 2000. "Effects of Greenness on Children's Cognitive Functioning." *Environment and Behavior* 32 (6): 775–795. doi:10.1177/00139160021972793

Wells, N. M., and G. W. Evans. 2003. "Nearby Nature: A Buffer of Life Stress among Rural Children." *Environment and Behavior* 35 (3): 311–330. doi:10.1177/0013916503035003001

Wong, H. K., and R. T. Wong. 2005. *The First Days of School: How to Be an Effective Teacher*. Mountain View, CA: Harry K. Wong Publications, Inc.

Woodhouse, J. L., and C. Knapp. 2000. *Place-Based Curriculum and Instruction: Outdoor and Environmental Education Approaches*. Clearinghouse on Rural Education and Small Schools, Appalachia Educational Laboratory.