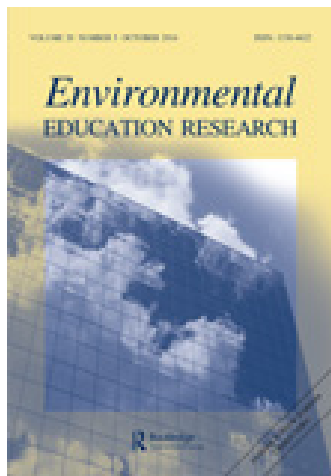


This article was downloaded by: [Virginia Tech Libraries]

On: 10 October 2014, At: 05:40

Publisher: Routledge

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Environmental Education Research

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/ceer20>

Environmental education program evaluation in the new millennium: what do we measure and what have we learned?

Marc J. Stern^a, Robert B. Powell^b & Dawn Hill^c

^a Department of Forest Resources and Environmental Conservation, Virginia Tech, Blacksburg, VA, USA

^b Parks, Recreation, and Tourism Management, Clemson University, Clemson, SC, USA

^c Psychology Department, University of Arizona South, Sierra Vista, AZ, USA

Published online: 23 Sep 2013.

To cite this article: Marc J. Stern, Robert B. Powell & Dawn 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](https://doi.org/10.1080/13504622.2013.838749)

To link to this article: <http://dx.doi.org/10.1080/13504622.2013.838749>

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Versions of published Taylor & Francis and Routledge Open articles and Taylor & Francis and Routledge Open Select articles posted to institutional or subject repositories or any other third-party website are without warranty from Taylor & Francis of any kind, either expressed or implied, including, but not limited to, warranties of merchantability, fitness for a particular purpose, or non-infringement. Any opinions and views expressed in this article are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor & Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Terms & Conditions of access and use can be found at <http://www.tandfonline.com/page/terms-and-conditions>

It is essential that you check the license status of any given Open and Open Select article to confirm conditions of access and use.

Environmental education program evaluation in the new millennium: what do we measure and what have we learned?

Marc J. Stern^{a*}, Robert B. Powell^b and Dawn Hill^c

^a*Department of Forest Resources and Environmental Conservation, Virginia Tech, Blacksburg, VA, USA;* ^b*Parks, Recreation, and Tourism Management, Clemson University, Clemson, SC, USA;* ^c*Psychology Department, University of Arizona South, Sierra Vista, AZ, USA*

(Received 10 December 2012; accepted 18 August 2013)

We conducted a systematic literature review of peer-reviewed research studies published between 1999 and 2010 that empirically evaluated the outcomes of environmental education (EE) programs for youth (ages 18 and younger) in an attempt to address the following objectives: (1) to seek reported empirical evidence for what works (or does not) in EE programming and (2) to uncover lessons regarding promising approaches for future EE initiatives and their evaluation. While the review generally supports consensus-based best practices, such as those published in the North American Association for Environmental Education's *Guidelines for Excellence*, we also identified additional themes that may drive positive outcomes, including the provision of holistic experiences and the characteristics and delivery styles of environmental educators. Overall, the evidence in support of these themes contained in the 66 articles reviewed is mostly circumstantial. Few studies attempted to empirically isolate the characteristics of programs responsible for measured outcomes. We discuss general trends in research design and the associated implications for future research and EE programming.

Keywords: environmental education; evaluation; meta-analysis; research; review

Introduction

Environmental education (EE) programs and organizations worldwide advance a wide range of goals believed to contribute to enhancing the environmental literacy of participants. This form of literacy is generally comprised of knowledge, attitudes, dispositions, and competencies believed to equip people with what they need to effectively analyze and address important environmental problems (Hollweg et al. 2011). The field of EE has also developed consensus-based guidelines for how to achieve this general goal, summarized most comprehensively in the North American Association for Environmental Education's (NAAEE) *Guidelines for Excellence* publications (NAAEE 2012a). These guidelines are based on the opinions of hundreds of researchers, theorists, and practitioners about what works in EE. As such, they provide lists and explanations of what might be considered generally agreed upon 'best practices' in the field.

*Corresponding author. Email: mjstern@vt.edu

This paper describes an effort to examine empirical evidence associated with those best practices. We conducted a systematic literature review of research published between 1999 and 2010 that empirically evaluated the outcomes of EE programs in an attempt to address the following objectives:

- To seek empirical evidence for what works (or does not) in EE programming.
- To identify lessons regarding promising approaches for future EE initiatives and their evaluation.

Systematic reviews of the empirical literature are not unprecedented in the EE literature. A number of such reviews have focused on program outcomes. For example, Leeming and his colleagues (1993) examined 34 EE studies between 1974 and 1991 that assessed knowledge, attitudes, and behavior-related outcomes of participants in in-class and out-of-class experiences. Zelezny (1999) analyzed 18 studies with a focus on whether programs in classroom or nontraditional settings performed better at improving environmental behavior. Each found significant shortcomings in the methods of the studies they reviewed. As such, they drew limited conclusions about the characteristics driving program outcomes. Rickinson (2001) identified 19 studies of EE programs between 1993 and 1999 that examined learning outcomes. He concluded from seven of those studies (including Zelezny's review) that certain program characteristics appear to facilitate positive outcomes. These characteristics included role modeling, direct experiences in the outdoors, collaborative group discussion, longer duration, and preparation and/or follow-up work. Rickinson explicitly noted a gap between speculation and evidence, however, and that empirical understanding of the characteristics which drive program outcomes 'is in need of further development.' (272)

Temporally, the current review starts where Rickinson left off, reviewing published articles since 1999. In contrast to prior reviews, we explicitly seek to examine the body of empirical evidence associated with consensus-based best practices in EE. We follow Rickinson's lead in an attempt to be systematic, comprehensive, and analytical. We developed clear criteria for inclusion of studies in the review and a common framework through which to analyze them. We aimed to include all peer-reviewed studies that met these criteria, regardless of our assessment of their individual quality. In this way, we could critically analyze and interpret the meaning of the findings reported therein and draw meaningful conclusions.

Consensus-based best practices in environmental education

Table 1 contains definitions developed from numerous sources that outline what many consider to be state-of-the-art, or 'best,' practices in EE (based primarily on EECO 2012; Hungerford, Bluhm, et al. 2001; Hungerford and Volk 1990; Hungerford, Volk, et al. 2003; Jacobson, McDuff, and Monroe 2006; Louv 2005; NAAEE 2012a). We focus most heavily on the NAAEE guidelines series (NAAEE 2012a), which was explicitly developed through consensus-building efforts amongst researchers and practitioners in EE. We used the definitions in Table 1 to track the characteristics of each program evaluated within the reviewed articles. We also tracked three additional practices that are generally less favored in the EE literature. These include: (1) traditional classroom approaches (Trad) in which the program consisted solely of traditional lecture style presentation(s); (2) lecture (Lect), in

Table 1. Definitions of program characteristics associated with the best practices.

Program Characteristic	Definition
Active participation (AP)	Participants are actively involved in the education experience, not just passive receivers of verbal or visual information or communication
Hands-on observation and discovery (HO)	Participants have to physically manipulate some aspect of the environment in some way to explore a concept or solve a problem
Place-based learning (Place)	The educational program is grounded in the particular attributes of a place, using <i>in situ</i> natural and community systems and themes as the context for learning. The content and learning is at least in some way <i>about</i> the specific place
Project-based learning (Proj)	Students are engaged in selecting, planning, implementing, and evaluating a real-world environmental project and making informed choices for action
Cooperative/group learning (C/G)	The learning environment requires participants to work with others, either through group deliberation/discussion and/or active participation/ investigation
Play-based learning (Play)	Students are actively engaged in games or competition as an intentional teaching technique
Outdoor instruction (Outside)	Instruction takes place outdoors
Investigation (Inv)	Students take part in data collection and/or basic inquiry
Guided inquiry (GI)	Educators ask questions and facilitate students' pursuit of answers
Pure inquiry (PI)	Participants develop their own research question(s), design and carry out research techniques to address that research question
Data collection (DC)	Scientific collection of data takes place (in any location)
Immersive field investigation (Imm)	Students are immersed in collecting real-world data <i>in the field</i> (a subcategory of 'Data Collection')
Relevance (Rel)	Content explicitly references or makes connections to students' experiences outside the realm of the instruction
Reflection (Ref)	Providing explicit opportunities for students to reflect on their prior experiences or new shared experiences (e.g. journaling or discussion)
Issue-based learning (Issue)	Curriculum focused on real-world environmental problems, their consequences, and potential solutions
Learner-centered instruction (LC)	Students control their own learning instead of having to follow formal curricula
Multimodal delivery of content (Mod)	Content is delivered in a way that uses more than one mode or media, requiring the use of other senses such as touch, smell, visual stimuli, etc
Multiple points of view (View)	Program explicitly acknowledges multiple points of view

which the program contained at least one lecture style presentation; and (3) programs that took place exclusively indoors (Inside). The abbreviations above and in Table 1 are used in subsequent tables to ease formatting.

While other terminology is common in the EE literature, we chose to focus upon practices we could clearly define and observe in the empirical literature. For example, while ‘constructivism’ is a ubiquitous term throughout the theoretical and empirical literature (e.g. Stern, Powell, and Ardoin 2010; Wright 2008; Yager 1991), it can manifest in multiple forms. Constructivist approaches help learners to construct their own understandings through building upon their prior knowledge and/or actively engaging them in real-world experiences (Jacobson, McDuff, and Monroe 2006). A program can be constructivist by directly relating content to participants’ prior experiences or home lives. Alternatively, programs can build new knowledge through shared experiential learning and enhance the constructivist nature of that knowledge through periodic reflection. As such, the term ‘constructivist,’ though commonly evoked in the reviewed studies, does not provide a single concrete practice in EE (e.g. DiEnno and Hilton 2005). In this case, we opted for more specific program elements that reflect different forms of constructivist program design – for example, explicit attempts for connecting program content to participants’ home lives, experiential learning, reflection, and field investigation. We also chose to break ‘experiential education’ into its many subcomponents, including active participation, hands-on observation and discovery, investigation, and other techniques that might commonly be considered ‘experiential’ (e.g. Jacobson, McDuff, and Monroe 2006).

Methodological approach

Our review sought empirical evidence of outcomes associated with each of the practices in Table 1 by identifying and examining peer-reviewed articles published since 1999 through 2010 that met the following criteria:

- A description of the EE program was included that was sufficient enough to identify the presence of program characteristics associated with the identified best practices.
- The program served youth audiences, ages 18 and below.
- There was clear evidence that subjects partook in the specific EE program.
- At least one element of knowledge, awareness, skills, attitudes, intentions, or behavior was empirically measured, either qualitatively or quantitatively, following participants’ exposure to the program.

Once identified, articles were reviewed and coded to achieve the following objectives:

- (1) To identify all relevant program characteristics of the evaluated programs that could be discerned from the article.
- (2) To identify all outcomes that were measured in each article and to assess the extent to which each outcome was generally positive or not.
- (3) To examine relationships between reported program characteristics and outcomes.
- (4) To identify authors’ explanations of results, in particular, those pertaining to which characteristics of the programs (or lack thereof) contributed most to the observed evaluation results.

- (5) To identify which of these explanations were speculative and which involved at least some degree of empirical evidence associated with the particular program characteristic's relationship to a desired outcome.
- (6) To summarize the research design and methods employed in each study and to identify and summarize shortcomings in research design or methods described by the authors.

Articles reviewed

To locate articles for the review, we began by reviewing the Table of Contents in mainstream EE journals published between 1999 and 2010, including *The Journal of Environmental Education*, *Environmental Education Research*, *Applied Environmental Education and Communication*, *International Research in Geographical and Environmental Education*, *Australian Journal of Environmental Education*, and the *Canadian Journal of Environmental Education*. We also conducted keyword searches in databases, including EBSCO Host® and Web of Knowledge®, and in other journals in which EE articles might typically appear, including *Environment and Behavior*, *International Journal of Science Education*, *Science Education*, and *Science Teacher*. Keywords included 'environmental education' and 'evaluation.' We also reviewed web-based repositories of EE studies, including NAAEE (2012b), PEEC (2012), and Zint (2012). We then reviewed the abstracts of all candidate articles whose titles revealed a likelihood of meeting the study criteria. As articles were identified, we scoured their literature-cited sections for other candidate articles and also conducted forward searches for articles that cited the articles already included in the review. We limited our final selection of articles to those focusing on school-aged youth (ages 18 and below), following Rickinson (2001). The search resulted in 66 articles that fully met the study criteria. Within these 66 articles, 86 EE programs (or groups of programs) were sufficiently described to be included in the analysis.

Article coding and analysis

At least two authors read each article and consulted on coding. The lead author was involved in the coding of all of the articles, while each other author read a majority. In cases of ambiguity or initial disagreement on coding, all three authors read and discussed the article. We came to consensus on the coding of both program characteristics (Table 1) and outcomes prior to conducting further analyses. For each of the 86 programs, we recorded the presence of each program characteristic described in the article. We also coded which of six outcomes of interest were measured in the article and the extent to which those outcomes were generally positive. Most articles assessed more than one outcome. To be able to provide consistent metrics across all articles, we settled upon the following outcomes' definitions for coding purposes:

- *Knowledge*: individual participants' change in knowledge of the subject after exposure to EE.
- *Awareness*: individual participants' change in recognition or cognizance of issues or concepts.
- *Skills*: individual participants' change in abilities to perform a particular action.

- *Attitudes*: individual participants' change in attitude toward the subject of the EE or environmental actions related to the programming.
- *Intentions*: individual participants' self-reported intent to change a behavior.
- *Behavior*: individual participants' self-reported behavior change or staff observations of behavior change following exposure to EE.
- *Enjoyment*: individual participants' overall satisfaction or enjoyment levels associated with the educational experience.

While other more nuanced outcomes were measured within the sample, for example, different types of knowledge, attitudes, skills, intentions, and behaviors, we opted for broader categories of outcomes to discern lessons that might crosscut the programs under study. These general categories also mirror the efforts of prior systematic reviews (Leeming et al. 1993; Rickinson 2001; Schneider and Cheslock 2003; Zelezny 1999). In an effort to be as inclusive as possible, the specific operational definitions listed above incorporate definitions from prior reviews and were finalized following our first reading of all articles in the review sample.

Other outcomes were occasionally observed, but not included in our analysis; the most common involved elements of empowerment, which we observed in eight studies. In three studies, this involved measuring of locus of control (Culen and Volk 2000; Dimopoulos, Paraskevopoulos, and Pantis 2008; Zint et al. 2002); in three other studies, self-efficacy was measured (Johnson-Pynn and Johnson 2005; Stern, Powell, and Ardoin 2010; Volk and Cheak 2003); and in two cases, more general forms of self-confidence were measured (Kusmawan et al. 2009; Russell 2000). We excluded 'empowerment' as an outcome in our analysis for three reasons: (1) its uncommon appearance and inconsistent conceptualization in the sample and (2) it was more commonly referenced in the reviewed articles as a probable reason for achieving another outcome than as an outcome itself. No other outcomes were common in the review that were not easily encompassed by the definitions provided above.

Consistent coding of outcomes as positive or negative presented perhaps the greatest challenge of the review, primarily due to the wide-ranging methodological paradigms employed within the reviewed studies. Following Rickinson (2001), we made a conscious effort to assess outcomes from within the research paradigm employed by the study authors. For example, the coding of outcomes for studies employing inferential statistics was based on tests of statistical significance. In these cases, however, ambiguities would arise when a study showed statistically significant results in some, but not all, measures of a particular outcome (e.g. two out of four measured outcomes showed positive change). Qualitative studies and those using only descriptive statistics posed even greater challenges for coding. To illustrate this ambiguity, consider the following: if evidence is presented that a program completely changed the life of one student, but not thirty others, should this be considered a 'positive' outcome? If data suggest that a program changed the attitudes of 40% of participants, should that be considered a 'positive' outcome? To reflect such ambiguities, we coded the data as follows, and analyzed it in two different ways.

Each analysis involved matching the presence of particular program characteristics with measured program outcomes. In our first analysis of the linkages between program characteristics and outcomes, we considered *any* positive finding for an outcome (including mixed measures for a single outcome) to be positive, erring on the side that any positive results, even if small or for a small proportion of students,

is a success. Using this definition, we witnessed very little variation in program outcomes across all studies. We present the first analysis in an appendix to this article (Table A). Because of the limited ability to observe variation in the first analysis, we performed a more nuanced second analysis. In the second analysis, presented in the manuscript, each outcome type (e.g. knowledge, attitude, intentions, etc.) within each article was coded as 'null,' 'mixed', or 'positive,' using the following definitions.

0 = Null (or negative) findings. In quantitative studies using inferential statistics, the article's author(s) report(s) no statistically significant positive result for the outcomes of interest. In qualitative studies, the author(s) explicitly note(s) a lack of positive outcomes. Note: only one study found a negative change in any outcome measures. (Martin et al. 2009)

1 = Mixed (or ambiguous) findings. In quantitative studies with inferential statistics, statistically significant positive changes are found for some, but not all, measures of a particular outcome. In studies using descriptive statistics only, less than 50%, but greater than 0%, of program participants exhibited positive outcomes. In qualitative studies, the author(s) explicitly note(s) that results for a particular outcome were mixed, or that some outcomes were positive while others were not. This category also includes mentions of positive outcomes for only some participants.

2 = Positive findings. In quantitative studies using inferential statistics, all measures of a particular outcome type exhibited statistically significant positive results. In studies using descriptive statistics only, at least 50% of program participants exhibited positive outcomes. In qualitative studies, the author(s) claim(s) only positive results (any null or conflicting cases are explicitly reported as exceptions or outliers).

We considered 50% a reasonable cut-off for descriptive studies, as we feel that if positive changes occur for a majority of participants, the program had an overall positive effect. We examined the relationships between the presence of each program characteristic and each outcome measure. To ease reporting, a mean score was calculated for each pairing of a program characteristic with each outcome type, such that a score of '0' would mean that the program characteristic was never associated with any positive outcome; a score of '1' would indicate mixed or ambiguous findings; and a score of '2' would signify that the program characteristic was only associated with unambiguous positive outcomes (as defined above).

To further explore linkages between program characteristics and measured outcomes, we also examined authors' claims about which aspects of the programs they felt were most important in driving study results. In these cases, we conducted qualitative coding on each article, focusing most energy on the discussion and conclusion sections. While we began this review with some pre-conceived categories, we allowed additional themes to emerge and felt it appropriate to focus on the language used by the authors in each case. This led to the development of themes that differ slightly from our preconceived list of 'consensus-based best practices.' We then examined each claim we recorded to determine whether any evidence existed within the article to empirically isolate the impact of the particular program characteristic upon a particular outcome. This evidence, when present, took multiple forms. In some cases, a control or comparison group provided evidence of enhanced outcomes in the presence of the program characteristic. In other cases, particularly in qualitative studies, study subjects, typically program participants (though in a few

cases parents or teachers), explicitly identified the aspects of the program that they felt led to a specific outcome. In keeping with our aim to analyze the data from within the particular paradigm of the researchers, we considered each as evidence of empirical isolation. We distinguish between inferential, descriptive, and qualitative evidence in the presentation of results.

Results

Programs

Forty-nine of the 86 evaluated programs included in the review took place in the USA. Programs varied tremendously in their duration, setting, and general nature. Sixteen were multiday residential experiences; 43 involved shorter field trips; 53 involved at least some time in the classroom; and 33 involved multiple settings. Thirty-five of the programs involved students aged 15–18, and 67 involved younger participants (4–14). Sixteen programs spanned both age groups.

Methods used in the studies

Fifty-seven studies employed quantitative techniques, and fifty-four employed qualitative techniques. Thirty-five were mixed methods studies. The most common study design was quasi-experimental, comparing pre-experience and post-experience student survey responses (50 studies). Twelve studies only measured outcomes after the experience. Fourteen studies involved a follow-up measure after some time had elapsed after the experience. Thirty-three studies employed some form of control or comparison group.

Authors of 12 studies suspected that particular weaknesses in their methods may have contributed to null findings. In most cases, authors felt that their measurements were not sensitive enough to detect changes between pre-experience and post-experience scores. In five cases, authors attributed this to ‘ceiling effects’ created by high scores prior to the experience. Other common measurement concerns included small sample sizes, vaguely worded survey items, unaccounted-for confounding factors, and social desirability bias (the case in which respondents select the answer they feel the surveyor is seeking, rather than that reflecting their true feelings).

Outcomes

Table 2 displays the number of programs for which each outcome of interest was measured and the percentage of times each was associated with a positive, mixed, or null result attributed to the program. Average outcomes scores are also presented (0–2 scale), along with the particular articles in which each outcome measure was observed. Knowledge was the most commonly measured outcome in the review, followed by attitudes. Null findings were not commonly reported within the sample. Only six programs were associated with null findings across all of their measured outcomes.

Program characteristics’ associations with outcomes

Program characteristics’ associations with outcomes were examined in three ways. We first examined the number of incidences in which selected program

characteristics were associated with *any* positive outcomes (see Appendix A, Table A). We then examined the data by combining all outcomes into a single measure and accounting for mixed results. As described in more detail above, outcomes were coded as ‘null’ if no measured outcomes were positive; ‘mixed’ if some measures were positive and others were null or mixed; and ‘entirely positive’ if all outcome measures were positive (Table 3). Finally, we examined the data associated with each outcome type separately while accounting for mixed results (Table 4). In this analysis, each outcome type (i.e. knowledge, skills, attitudes, intention, behavior, and enjoyment) is considered on its own. The table is sorted by a weighted average that accounts for each individual pairing of each program characteristic and each outcome. Awareness is not included in the table because only one program displayed null findings for awareness. That program employed a traditional, lecture-only approach indoors.

The tables, taken together, provide circumstantial evidence in favor of all the consensus-based best practices observed in the literature. The five traditional programs, typically included as control groups in the reviewed studies, clearly achieved less positive outcomes than other programs. While some practices appear to be somewhat more commonly associated with better outcomes than others (for example, investigation, projects, and reflection), the general low variability in outcomes does little to differentiate the strength of each practice in a practical sense. Moreover, the coincidence of program characteristics with positive, mixed, or null outcomes does not mean the two are necessarily causally linked. Our review cannot account for the quality with which each practice was performed, nor could we account for additional characteristics that may have been present in the program but not described by the author(s). An additional limitation involves the enhanced likelihood of achieving what we have termed ‘mixed’ results in studies in which a larger number of outcomes were measured. To address these limitations, we re-examined each article to explore authors’ explanations for the program outcomes they observed.

Reviewed studies’ authors’ explanations for observed outcomes

Arguments posited by the authors of the reviewed studies for why (or why not) the observed programs achieved particular outcomes are summarized in Table 5. The list is somewhat different from the specific practices contained in the earlier analysis, as authors often used broader terms to encompass a suite of practices. For example, experiential education encompasses descriptions of students’ active engagement in first-hand experiences. This includes multiple versions of active, hands-on learning. In other cases, authors posited explanations that were not identified in our literature review of best practices – for example, the style or identity of the educator.

Table 6 displays the number of studies containing each claim, the outcomes with which claims were associated, and the frequency with which some evidence was actually present to empirically support the authors’ speculation about a particular program characteristic. In Table 6, we distinguish between three types of empirical support: (1) inferential statistics (I); (2) qualitative findings accompanied by descriptive statistics (D); and (3) qualitative findings without descriptive statistics (Q). In inferential cases, evidence involved a control or comparison group which contributed to isolating the role of a characteristic in influencing outcomes (e.g. Basile 2000; Liu and Kaplan 2006; Siemer and Knuth 2001; Stern, Powell, and Ardoin 2008). In other

Table 2. Outcomes measured across all evaluated programs.

Outcome	Programs in which outcome was measured	Positive findings			Average outcome score	Articles
		Null findings (%)	Mixed (%)	Positive (%)		
Knowledge	76	8	11	82	1.74	Agelidou, Balafoutas, and Flogaitis (2000), Aivazidis, Lazaridou, and Hellden (2006), Ajiboye and Olatundun (2010), Andrews, Tressler, and Mintzes (2008), Ballantyne and Packer (2002), Ballantyne and Packer (2009), Ballantyne, Fien, and Packer (2000), Ballantyne, Fien, and Packer (2001), Basile (2000), Baumgartner and Zabin (2008), Bexell et al. (2009), Bodzin (2008), Bogner (1999); Bradley, Waliczek, and Zajicek (1999), Braun, Buyer, and Randler (2010), Cachelin, Paisley, and Blanchard (2009), Carrier (2009), Cullen and Volk (2000), Cummins and Snively (2000), D'Agostino et al. (2007), DiEonno and Hilton (2005), Dimopoulos, Paraskevopoulos, and Pantis (2008), Ernst (2005), Farmer, Knapp, and Benton (2007), Flowers (2010), Gambino, Davis, and Rowntree (2009), Grodzinska-Jureczak et al. (2003), Hansel et al. (2010), Johnson-Pynn and Johnson (2005), Klosterman and Sadler (2010), Knapp and Benton (2006), Knapp and Poff (2001), Kruse and Card (2004), Kuhar et al. (2007), Kusmawan et al. (2009), Lindemann-Matthies (2002), Liu and Kaplan (2006), Malandrakis (2006), Mayer-Smith, Bartosh, and Peterat (2009), Middlestadt et al. (2001), Miller (2007), Morgan et al. (2009), Poudel et al. (2005), Powell and Wells (2002), Powers (2004), Randler, Ilg, and Kern (2005), Ruiz-Mallen et al. (2009), Russell (2000), Salata and Ostergren (2010), Siemer and Knuth (2001), Stern, Powell, and Ardoin (2008), Vaughan et al. (2003), Villegas et al. (2010); Volk and Cheak (2003), and Zint et al. (2002)
Awareness	17	6	0	94	1.88	Ballantyne, Fien, and Packer (2001), Cullen and Volk (2000), Gastreich (2002), Grodzinska-Jureczak et al. (2003), Johnson-Pynn and Johnson (2005), Kusmawan et al. (2009), Miller (2007), Morgan et al. (2009), Ruiz-Mallen et al. (2009), Salata and Ostergren (2010), Schneller (2008), Smith-Sebasto and Semrau (2004), Stern, Powell, and Ardoin (2008), and Stern, Powell, and Ardoin (2010)
Skills	25	8	24	68	1.60	Agelidou, Balafoutas, and Flogaitis (2000), Ballantyne, Fien, and Packer (2000), Basile (2000), Baumgartner and Zabin (2008), Cullen and Volk (2000), Ernst and Monroe (2004), Ernst (2005), Flowers (2010), Gambino, Davis, and Rowntree (2009), Miller (2007), Morgan et al. (2009), Nicolaou et al. (2009), Russell (2000), Siemer and Knuth (2001), Smith-Sebasto and Semrau (2004), Stern, Powell, and Ardoin (2010), Volk and Cheak (2003), and Zint et al. (2002)

(Continued)

Table 2. (Continued).

Outcome	Programs in which outcome was measured	Null findings (%)		Positive findings		Average outcome score	Articles
				Mixed (%)	Positive (%)		
Attitudes	54	26	37	37	37	1.11	Aivazidis, Lazaridou, and Hellden (2006), Andrews, Tressler, and Mintzes (2008), Ballantyne and Packer (2002), Ballantyne and Packer (2009), Ballantyne, Fien, and Packer (2000), Ballantyne, Fien, and Packer (2001), Bexell et al. (2009), Bodzin (2008), Bogner (1999), Bradley, Waliczek, and Zajicek (1999), Braun, Buyer, and Randler (2010), Cachelin, Paisley, and Blanchard (2009), Carrier (2009), Culen and Volk (2000), Cummins and Snively (2000), Dettmann-Easler and Pease (1999), DiEnno and Hilton (2005), Dimopoulos, Paraskevopoulos, and Pantis (2008), Eagles and Demare (1999), Ernst (2005), Farmer, Knapp, and Benton (2007), Flowers (2010), Gambino, Davis, and Rowntree (2009), Johnson and Manoli (2008), Knapp and Poff (2001), Kruse and Card (2004), Kuhar et al. (2007), Kusmawan et al. (2009), Liu and Kaplan (2006), Malandrakis (2006), Martin et al. (2009), Mayer-Smith, Bartosh, and Peterat (2009), Middlestadt et al. (2001), Nicolaou et al. (2009), Powers (2004), Schneller (2008), Siemer and Knuth (2001), Smith-Sebasto and Cavern (2006), Smith-Sebasto and Semrau (2004), Stern, Powell, and Ardoin (2008), and Stern, Powell, and Ardoin (2010)
Intentions	23	26	26	26	48	1.22	Ballantyne and Packer (2002), Ballantyne and Packer (2009), Ballantyne, Fien, and Packer (2000), Ballantyne, Fien, and Packer (2001), Bexell et al. (2009), Bogner (1999), Dimopoulos, Paraskevopoulos, and Pantis (2008), Flowers (2010), Johnson-Pynn and Johnson (2005), Knapp and Poff (2001), Stern, Powell, and Ardoin (2008), Stern, Powell, and Ardoin (2010), and Zint et al. (2002)
Behavior	19	11	74	16	16	1.05	Ballantyne, Fien, and Packer (2000), Baumgartner and Zabin (2008), Bexell et al. (2009), Carrier (2009), Culen and Volk (2000), Gambino, Davis, and Rowntree (2009), Gastreich (2002), Grodzinska-Jurezak et al. (2003), Knapp and Poff (2001), Kruse and Card (2004), Middlestadt et al. (2001), Schneller (2008), Siemer and Knuth (2001), Smith-Sebasto and Semrau (2004), Stern, Powell, and Ardoin (2008), Stern, Powell, and Ardoin (2010), and Volk and Cheak (2003)
Enjoyment	21	10	10	80	80	1.71	Ballantyne and Packer (2002), Ballantyne, Fien, and Packer (2001), Bogner (1999), Braun, Buyer, and Randler (2010), DiEnno and Hilton (2005), Grodzinska-Jurezak et al. (2003), Knapp and Poff (2001), Morgan et al. (2009), Ruiz-Mallen et al. (2009), Russell (2000), Salata and Ostergren (2010), and Stern, Powell, and Ardoin (2010)

cases, qualitative data were presented to directly support the role of a particular practice in influencing student outcomes. This commonly took the form of interview data in which participants answered questions about the aspects of programs that led to their own perceived changes in knowledge, attitudes, or behavioral measures (e.g. Ballantyne, Fien, and Packer 2000; Ballantyne and Packer 2002; Ballantyne and Packer 2009; Knapp and Poff 2001; Russell 2000). In a few cases, a particular practice was associated with mixed outcomes – for example, two out of three attitude measures showed a positive response. These cases are marked by bold italics in the table. Seventeen studies provided some form of empirical isolation of a program characteristic's relationship with a measured outcome. Ten articles used inferential statistics; five used descriptive statistics; eight provided un-quantified qualitative evidence; and six articles provided more than one form of evidence.

Experiential education was the most commonly hypothesized explanation for a program's degree of success, followed by issue-based education, direct contact with nature, dosage, investigation, and empowerment. Few of the claims made by authors were directly supported with empirical evidence. The most common empirically supported claims involved experiential education, dosage, and investigation.

While the presence of empirical support certainly bolsters authors' claims, the lack of it does not necessarily mean that a particular claim may be any less valid. For example, claims associated with participants' development of emotional connections may be just as important as other claims even though they were not systematically measured. In many cases, authors made claims based on their own detailed observations. In other cases, however, claims were clearly made based upon pre-existing theory or general impressions of what might have made a program more successful in the absence of any real data. In most of these cases, the authors were clear about the speculative nature of their propositions.

In some cases, empirical support existed for a suite of practices, rather than any particular single program characteristic. For example, DiEnno and Hilton (2005) isolated the impact of what they refer to as a 'constructivist' approach upon participants' knowledge and attitudes. In this case, the elements of the program that differed from a control group included cooperative/group learning, place-based, and issue-based education. Kusmawan and colleagues (2009) provided empirical support for programs that incorporated both field-based investigations and student engagement with local community members in local environmental issues. While general claims about experiential education were supported in the article, it was unclear which particular aspects of each program were responsible for driving specific differences in outcomes. This highlights both the methodological challenges of isolating best practices and the interacting components of programs that contribute to a holistic experience responsible for overall outcomes in participants. Groupings of multiple characteristics as clusters are not included in Table 6, with the exception of the IEEIA model, which is an explicitly described approach in each case.

Interpretation and discussion

While unable to conclusively isolate the key characteristics that tend to produce the most desired outcomes in EE, this review provides some basic, though inconclusive, evidence in support of the existing consensus-based guidelines for EE programming. Authors commonly highlighted certain program characteristics in particular (Table 5). Authors' claims may be interpreted in at least two ways, however. As observers

Table 3. Coincidence of program characteristics with combined outcomes findings.

Characteristic	N	Null findings (%)	Positive findings		Average outcome score (0–2 scale)
			Mixed (%)	Entirely positive (%)	
Immersive field investigation (Imm)	15	0	33	67	1.67
Data collection (DC)	27	0	41	59	1.59
Guided inquiry (GI)	17	0	41	59	1.59
Project-based learning (Proj)	24	0	42	58	1.58
Investigation (Inv)	40	0	48	53	1.53
Reflection (Ref)	35	3	43	54	1.51
Hands-on observation & discovery (HO)	58	3	47	49	1.47
Active participation (AP)	70	4	46	49	1.46
Cooperative/group learning (C/G)	31	0	58	42	1.42
Multimodal delivery of content (Mod)	61	7	46	48	1.41
Issue-based learning (Issue)	42	2	55	43	1.40
Outdoor instruction (Outside)	56	4	54	43	1.39
Learner-centered instruction (LC)	11	0	64	36	1.36
Inside only (Inside)	22	9	45	45	1.36
Place-based learning (Place)	52	4	58	38	1.35
Play-based learning (Play)	13	8	54	38	1.31
Relevance (Rel)	23	4	61	35	1.30
Lecture (Lect)	25	8	56	36	1.28
Pure inquiry (PI)	4	0	75	25	1.25
Multiple points of view (View)	9	0	78	22	1.22
Traditional (Trad)	5	40	60	0	0.60
Overall programs	86	7	50	43	1.36

Notes: Positive (2) = only positive findings on *all* measured outcomes within a study. Mixed (1) = some positive findings; some null findings. Null (0) = only null or negative findings.

Table 4. Outcome scores associated with observed program characteristics across the 66 articles in the review.

Reported practices	Desired outcomes							Weighted average
	Knowledge	Skills	Attitudes	Intention	Behavior	Enjoyment		
Imm	Average	1.93	1.89	1.57	1.67	1.40	2.00	1.79
	Observations	14	9	7	3	5	4	42
Proj	Average	2.00	1.58	1.55	1.50	1.33	1.83	1.70
	Observations	19	12	11	6	6	6	60
DC	Average	1.84	1.75	1.40	2.00	1.22	1.86	1.67
	Observations	25	8	15	5	9	7	69
Inv	Average	1.89	1.73	1.26	1.50	1.27	1.92	1.63
	Observations	35	15	23	10	11	12	106
Ref	Average	1.83	1.64	1.37	1.50	1.27	1.92	1.62
	Observations	29	11	19	6	9	7	81
PI	Average	2.00	1.67	1.00	2.00	1.67	1.00	1.62
	Observations	3	3	2	1	3	1	13
GI	Average	1.76	1.67	1.33	2.00	1.22	1.86	1.59
	Observations	17	3	9	1	5	2	37
HO	Average	1.85	1.65	1.19	1.44	1.19	1.92	1.57
	Observations	52	20	32	16	16	12	148
C/G	Average	1.93	1.70	1.16	1.29	1.20	1.60	1.55
	Observations	28	10	19	7	10	10	84
Issue	Average	1.86	1.53	1.31	1.43	1.25	1.57	1.55
	Observations	36	17	26	14	12	11	116
AP	Average	1.81	1.62	1.23	1.28	1.12	1.93	1.54
	Observations	63	24	39	18	17	14	175
Place	Average	1.84	1.61	1.25	1.35	1.07	1.80	1.53
	Observations	43	18	36	17	12	15	141
LC	Average	2.00	1.80	1.17	1.00	1.33	1.25	1.52
	Observations	7	5	6	2	3	4	27

(Continued)

Table 4. (Continued).

Reported practices	Desired outcomes							Weighted average
	Knowledge	Skills	Attitudes	Intention	Behavior	Enjoyment		
Mod	1.79	1.75	1.19	1.24	1.08	1.73		1.51
Observations	53	16	37	17	12	15		150
Outside	1.83	1.71	1.23	1.29	1.07	1.80		1.51
Observations	48	14	43	17	15	15		152
View	1.83	2.00	0.83	1.25	1.40	1.80		1.48
Observations	6	3	6	4	5	5		29
Rel	1.89	1.43	1.13	1.50	1.13	1.63		1.48
Observations	18	7	16	6	8	8		63
Lect	1.67	1.83	0.71	0.67	1.00	1.50		1.29
Observations	24	6	17	3	7	6		63
Inside	1.62	1.56	0.50	1.00	0.50	0.00		1.23
Average	21	9	10	6	2	1		49
Play	1.58	1.67	1.00	1.00	0.75	2.00		1.35
Observations	12	3	8	5	4	5		37
Trad	0.80	NA	0.00	0.00	0.00	0.50		0.36
Observations	5		5	1	1	2		14
Total	1.88	1.77	1.22	1.40	1.21	1.82		1.53
Observations	558	213	386	165	172	162		1656

Notes: Scores range from 0 to 2. Scores above 1 show on average at least some positive relationship between the characteristic and the associated outcome. Scores below 1 indicate more commonly null findings.

Table 5. Authors' speculative claims about reasons for program success or failure.

Author claim	Definition
Experiential	Learning was explicitly experiential. That is, students were engaged in actively participating in a particular firsthand experience
Issue-based	Education focused on specific environmental issues
Dosage	The duration and/or frequency of programs. Also, pre- and post-experience preparation and follow-up
Direct contact with nature	Program took place outside and/or included direct contact with living organisms
Empowerment	Program was observed to explicitly enhance perceptions of self-efficacy on the students
Investigation component	Program involved investigation
Place-based	Program was on-site and at least in some way <i>about</i> a particular place and students' connection to it
Explicit reflection	Program required students to explicitly reflect on their experiences
Project-based	Students took part in a specific project (such as a community project, a clean-up, or developing and delivering presentations to others)
Connection to home lives of participants	Content explicitly made connections to students' experiences outside the realm of the instruction
Affective elements incorporated	Author cited emotional connections as key to programs' success
Active participation	Students were more than just passive receivers of information
Focus on consequences of environmental problems	Consequences of environmental damage and/or solutions were explicitly demonstrated and discussed
Style or identity of educator	Author attributed program success to something about the specific personality or identity of the educator
Group discussion or cooperative group work	Participants worked with others, either through group deliberation/discussion and/or active participation/investigation
Mismatch between goals and program design	Author explained null findings by citing that the program had no design element one might expect would lead to the outcome
Multiple generations	Multiple generations were explicitly and actively included in a program

(Continued)

Table 5. (Continued).

Author claim	Definition
School teacher involvement	Students' school teachers were actively engaged as instructors or co-instructors within the EE program
IEEIA-based intervention	Program was explicitly based on the <i>Investigating and Evaluating Environmental Issues and Actions</i> framework (Hungerford, Volk, et al. 2003)
Multiple viewpoints	Multiple viewpoints about an issue were explicitly shared within the program
Play-based learning	Education explicitly incorporated play
Multimodal/sensory engagement	The program explicitly engaged multiple senses beyond the traditional senses used in education. For example, smell, additional visualizations, touch, taste, etc
Schematization of concepts	Education explicitly incorporated the development of causal chains or concept maps as a teaching/learning tool

Table 6. Summary of authors' claims about individual program characteristics and empirical support for those claims.

Explanatory factor	Empirical isolation										References
	Studies	Outcomes	Studies	KN	AW	SK	ATT	INT	BEH	ENJ	
Experiential	25	All	6	I(1), IQ(1), D(1), Q(3)	IQ(1)	I(1), D(1)	I(2), IQ(1), D(2)	I(1)	I(1), D(1)	I(1)	Ballantyne, Fien, and Packer (2000), Ballantyne and Packer (2009), Knapp and Poff (2001), Kusmawan et al. (2009), Russell (2000), and Steiner and Knuth (2001)
	23	All	3	D(1), Q(1)	Q(1)	D(1), Q(1)	D(1)	Q(1)	D(1)	D(1)	Ballantyne, Fien, and Packer (2000) and Ballantyne, Fien, and Packer (2001)
Direct contact with nature	23	All	3	D(2), Q(1)	D(1)	D(1)	D(3), Q(1)	D(1), Q(1)	D(2)	D(1)	Ballantyne, Fien, and Packer (2000), Ballantyne, Fien, and Packer (2001), Ballantyne and Packer (2002), and Ballantyne and Packer (2009)
Dosage	18	All	5	I(2)	I(1), Q(1)	I(2), Q(1)	I(1)	I(1)	I(1)	I(1)	Ballantyne and Packer (2002), Powers (2004), Ruiz-Mallen et al. (2009), Stern, Powell, and Ardoin (2008), and Smith-Sebasto and Cavem (2006)
Pre-experience and post-experience activities (dosage special case)	10	All	3	I(1)	I(1)	I(2)	I(2)	I(2)	I(2)	I(2)	Sebasto and Cavem (2006), Ballantyne and Packer (2002), Stern, Powell, and Ardoin (2008), and Smith-Sebasto and Cavem (2006)
Investigation component	16	All	5	D(3)	D(1)	I(1), D(1)	D(2)	D(3)	D(2)	Q, D(2)	Ballantyne, Fien, and Packer (2000), Ballantyne, Fien, and Packer (2001), Ballantyne and Packer (2002), Ballantyne and Packer (2009), and Basile (2000)
Empowerment (focus on self-efficacy and general confidence)	16	All	1					I(1)	Q(1)	Q(1)	Kusmawan et al. (2009)
Project-based	13	All	2	D(1), Q(2)		Q(1)	Q(1)	Q(1)	Q(1)	Q(1)	Ballantyne, Fien, and Packer (2000) and Russell (2000)
Place-based	12	All	3	Q(2)	D(1)	Q(1)	D(1), Q(2)	Q(1)	D(1), Q(1)	D(1)	Ballantyne, Fien, and Packer (2000), Ballantyne, Fien, and Packer (2001), and Ballantyne and Packer (2009)
Explicit reflection	11	All	1	D(1)		D(1)	D(1)	D(1)			Ballantyne and Packer (2009)
Active participation	11	All	–								Ballantyne and Packer (2009)

(Continued)

Table 6. (Continued).

Explanatory factor	Speculative claims				Empirical isolation						References
	Studies	Outcomes	Studies	KN	AW	SK	ATT	INT	BEH	ENJ	
Focus on consequences of environmental problems	10	All	3	D(1), Q(1), ID(1)		Q(1)	D(1), Q(1), ID(1)	D(1), ID(1)	D(1), Q(1), D(1), Q(1), ID(1)		Ballantyne, Fien, and Packer (2000), Ballantyne, Fien, and Packer (2001), and Ballantyne and Packer (2002)
Style or identity of educator/presenter	10	All except SK	–			Q(1)	Q(1)	Q(1)	Q(1)		Ballantyne, Fien, and Packer (2000)
Connection to home lives of participants	10	All except INT	1	Q(1)							Russell (2000)
Affective elements incorporated	9	All	–			Q(1)					Liu and Kaplan (2006)
Group discussion or cooperative group work	8	All	1								Russell (2000)
Program misalignment with outcomes measures	7	All	–				I(1)				Culen and Volk (2000) and Volk and Cheak (2003)
Multiple generations	5	All except SK	1	Q(1)					I(2)		Stern, Powell, and Ardoin (2008)
Play-based learning	5	All	1	I(1, /)							
IEEIA-based intervention	4	All	2	I(1, /)							
School teacher involvement	4	All except ENJ	1	I(1)	I(1)			I(1)	I(1)		Ageidou, Balaifoutas, and Flogaitis (2000)
Multiple viewpoints	4	All	–								
Multimodal/sensory engagement	4	KN, AW, SK, ATT	–								
Schematization of concepts	2	KN, SK	1	I(1)							

Notes: KN = knowledge; AW = awareness; SK = skills; ATT = attitudes; INT = intentions; BEH = behaviors. I = inferential statistics; D = qualitative with descriptive statistics; Q = qualitative without descriptive statistics. Bold italics signify mixed results (e.g. some, but not all measures of a construct were positive influenced).

(typically, but not always, external to the program itself) who are focused explicitly on program evaluation, they are commonly positioned well to draw valid conclusions about the most effective or ineffective aspects of the programs. However, most authors clearly espouse particular theoretical perspectives (at least in the introductory sections of their papers) that may predispose them to focus on certain programmatic elements at the expense of others. As such, synthesizing authors' claims about what worked or did not work in their programs is instructive both in terms of identifying the most promising practices (particularly in cases where empirical support exists) and further illuminating the dominant assumptions (and perhaps blind spots) of the field.

Insights on effective EE practices

The review suggests that a number of program elements may positively influence outcomes of EE programs. First, *active and experiential engagement in real-world environmental problems* appears to be in favor with EE researchers and empirically supported. In particular, *issue-based*, *project-based*, and *investigation-focused* programs in *real-world nature settings* (place-based) commonly achieved desired outcomes, and authors commonly attributed positive outcomes to these particular program attributes. The importance of *empowerment* and *student-centered learning* geared toward developing skills and perceptions of self-efficacy in these types of programs was also supported in this review. Further evidence of the promise of these practices exists in prior reviews of the IEEIA model, which aims to provide a holistic experience in which students investigate real-world environmental issues through a multidisciplinary approach that leads them to identify and deliberate appropriate courses of action (Hungerford and Volk 1990; Hungerford, Volk, and Ramsey 2000; Hungerford, Volk, et al. 2003). Two studies in the current review provided some empirical evidence in favor of this model as well (Culen and Volk 2000; Volk and Cheak 2003), though some outcome measures (attitudes and skills) were mixed in these cases.

Many authors also attributed success to various forms of *social engagement*. This most commonly took the form of *cooperative group work* amongst students. However, authors also noted particular value in involving *inter-generational communications* within a program and certain forms of *teacher engagement*. For example, one study found that when school teachers on field trips actively participated in the onsite instruction alongside EE instructors, students' outcomes were generally more positive (Stern, Powell, and Ardoin 2008). The findings suggest the importance of the roles of teachers and other adults as role models in developing environmental literacy (Emmons 1997; Rickinson 2001; Sivek 2002; Stern, Powell, and Ardoin 2008; Stern, Powell, and Ardoin 2010). An alternative interpretation may be that teacher's participation in the onsite instruction developed more fully their own meanings of the experience, which allowed them to reinforce student learning during and after the experience.

A number of authors noted the *identity and/or style of the instructor* as a primary driver of positive outcomes for students. Interestingly, this particular theme is not overtly present within the NAAEE guidelines, nor was it explicitly built into any research designs encountered in this review. The formal education literature, however, has long considered teachers' verbal and non-verbal communication styles to be prominent determinants of student outcomes (Finn et al. 2009). Moreover, a

recent study conducted by the first two authors of this article revealed that certain characteristics of educators (interpretive rangers in the National Park Service in this case), in particular their comfort, eloquence, apparent knowledge, passion, sincerity, and charisma, were strongly associated with more positive visitor outcomes. These outcomes included satisfaction with the program, enhanced appreciation of resources, and behavioral intentions (Stern and Powell, [forthcoming](#)). Authors' claims in the current literature review further support the importance of demonstrating *passion for the subject matter* and *genuine care and concern for students* (e.g. Ballantyne, Fien, and Packer 2001; Russell 2000), though explicit empirical tests pertaining to these characteristics were lacking.

The authors of nine studies suspected that the *emotional connections* made during their programs were the primary drivers of the measured outcomes. The emotional connections were made in multiple ways in the study sample, ranging from *interactions with animals and places* to *extensive group discussion* and *collaboration involving communities and real-world problems*. The importance of the affective domain has been discussed in the EE literature extensively (e.g. Iozzi 1989; Sobel 2012) and also mirrors guidance from the field of interpretation (Skibins, Powell, and Stern 2012; Stern and Powell, [forthcoming](#); Tilden 1957).

Some of the most successful programs in the review also highlight another element common in the interpretation field, but less commonly noted in the EE field: the concept of providing a *holistic experience* (Skibins, Powell, and Stern 2012; Stern and Powell, [forthcoming](#); Tilden 1957). Holistic experiences involve conveying a complete idea or story within the educational context. They thus carry high potential to provide a coherent picture of the relevance of the educational activity and a clear take-home point for students to reflect upon or pursue following the experience. This may often require *pre-experience preparation* and/or *post-experience follow-up* to an onsite educational experience. Smith-Sebasto and Cavern's study (2006) lends particular support to this idea in which neither pre-experience preparation nor post-experience follow-up on their own enhanced students' environmental attitudes. Only when both were present were gains witnessed in students' attitudes. Multiple studies evaluated programs in which students were placed within the story and asked to play an active role in learning about a problem or issue, investigating and evaluating that problem, and debating appropriate courses of action (e.g. Culen and Volk 2000; Volk and Cheak 2003). Other successful programs focused on *specific places and issues, explicitly linking program content to students' home lives*, and/or *explicitly provoking student reflection* (e.g. Ballantyne, Fien, and Packer 2000; Kusmawan et al. 2009; Stern, Powell, and Ardoin 2010). Similar to being able to step into the foreground of a landscape painting, such elements may allow students to step into the issue and recognize their connections to it. The review revealed each of these practices to be commonly associated with positive outcomes, and each was cited by authors as potential drivers of those outcomes.

Overall, consensus-based best practices in EE were broadly, though only circumstantially, supported in the literature review. However, additional concepts from interpretation and formal education associated with holistic experience-making, affective (emotional) messaging, and passionate, confident, caring, and sincere delivery also appear to be highly relevant to influencing EE program outcomes. While these elements are certainly not absent from the EE literature, they were not commonly part of any initial focus of the articles contained in this review. While multiple consensus-based best practices implicitly suggest the importance of holistic

experiences (e.g. Hungerford, Volk, et al. 2003; NAAEE 2012a), and the affective domain has long been identified as an important component of learning in the EE context (e.g. Iozzi 1989), their absence as central concepts in the empirical literature suggests a lack of clear focus on the importance of these particular elements in program design. Moreover, the lack of attention to the characteristics and delivery styles of educators in favor of content-based guidelines ignores whole bodies of literature within the fields of formal education and communication (e.g. Finn et al. 2009). As such, additional attention to these elements may be warranted within both EE research and practice.

Insights on EE evaluation

The review suggests a number of lessons for EE evaluation research and its potential role in supporting the enhancement of EE programming. We examined the most recent decade's published efforts in EE program evaluation in an attempt to further our understanding of how particular elements of EE programs contribute to variable outcomes. We found broad evidence that EE programs can lead to positive changes in student knowledge, awareness, skills, attentions, intentions, and behavior. However, we found only circumstantial evidence related to how or why these programs produce these results. We conclude that the current practice of EE program evaluation is, for the most part, not oriented toward studies that enable the empirical isolation and/or verification of particular practices that tend to most consistently produce desired outcomes.

In a recent review of EE evaluation research, Carleton-Hug and Hug (2010) noted that most published EE evaluation research represents utilization-focused evaluation (Patton 2008) and summative evaluation. Our review clearly reflects the same trend. Each of these approaches tends to focus on the unique characteristics and goals of individual programs. Utilization-focused evaluations, along with the emergence of participatory evaluation approaches (e.g. Powell, Stern, and Ardoin 2006), often develop unique measures of outcomes based on the goals of a particular program, limiting the direct comparability of outcomes across studies. This raises the question of whether these approaches constrain the ability of the field to unearth broader lessons that might be uncovered if a broader and more consistent suite of outcomes was commonly measured (e.g. Hollweg et al. 2011). Summative evaluations commonly forego opportunities to examine the influence of particular program elements upon measured outcomes (Carleton-Hug and Hug 2010). This may also be one explanation for the common lack of clear program description we observed in our review. Moreover, the focus on single programs inhibits the ability of researchers to understand the influence of context (Carleton-Hug and Hug 2010) unless explicitly built into the research design (e.g. Powers 2004; Stern, Powell, and Ardoin 2010).

The general dearth of null results in the review may also be related to the widespread practice of utilization-focused (Patton 2008), summative (Carleton-Hug and Hug 2010), and participatory (e.g. Powell, Stern, and Ardoin 2006) evaluation approaches. These approaches typically seek to align evaluation measures as closely as possible with programmatic goals and, as such, may be less likely to find null results. The lack of null results limits the prospects for meta-analyses in which clear patterns may be detected.

On the other hand, evidence emerged in our review that outcome measures are sometimes *not* directly related to program content. Seven authors attributed null findings to a mismatch between program content and measured outcomes. This is a well-known concern of EE and interpretation research, particularly with regard to influencing behavioral outcomes (Carleton-Hug and Hug 2010; Ham 2013; Monroe 2010). Decades of research on human behavior broadly recognize that knowledge gain is not typically a direct cause of behavior change (Ajzen 2001; Hines, Hungerford, and Tomera 1987; Hungerford and Volk 1990). As such, programs that focus primarily on providing new knowledge should not be expected to necessarily influence behavioral outcomes, even though they may measure them (Ham 2013; Jacobson, McDuff, and Monroe 2006; Stern and Powell, *forthcoming*). A similar theme holds for the tenuous relationship between attitudes and behaviors, especially in the environmental domain (Heimlich and Ardoin 2008; Jacobs et al. 2012). As noted by others (Heimlich 2010; Monroe 2010), researchers could potentially play a stronger role in articulating theories relevant to program design, not only to ensure appropriate measures, but perhaps more importantly to enhance program design and reformulation (e.g. Powell, Stern, and Ardoin 2006).

While long-standing definitions of EE typically incorporate some aspects of knowledge, EE goals typically stress influencing the behaviors of participants (Heimlich 2010; Hungerford and Volk 1990; UNESCO 1978). The prevalence of knowledge as the most commonly measured outcome across the studies in this review may suggest a number of trends. Are programs failing to target behavioral outcomes? Are standardized tests and/or other school curriculum requirements driving EE programs to focus more on knowledge provision or are educators choosing to do so? Are researchers falling short on measuring behavioral outcomes? Or is knowledge typically measured simply because it tends to be easier to operationalize than other potential outcomes? Regardless of the answers to these questions, this review suggests that knowledge gain remains a central focus of the EE evaluation field.

It is clear that the field could benefit from studies that aim to empirically isolate program components that tend to lead to more positive outcomes for students. Most current studies evaluate singular programs and thus are incapable of any more than speculation about why a particular program achieved its particular outcomes. Comparative studies, such as those undertaken by Ballantyne, Packer, and Fien (Ballantyne, Fien, and Packer 2000; Ballantyne, Fien, and Packer 2001; Ballantyne and Packer 2002; Ballantyne and Packer 2009), show particular promise in this sense. Ideally, a larger scale study, such as the one we undertook in the interpretive field that tracked similar program characteristics and outcomes over 376 programs (Stern and Powell, *forthcoming*), could be developed to further determine which approaches are most likely to be successful in varying conditions and with various audiences. In the absence of such a study, we encourage researchers performing program evaluations to focus on providing additional details of the programs they evaluate, using control or comparison groups wherever possible, and/or using retrospective qualitative interviews to contribute to the overall effort of understanding not only if EE works, but also by *why* and *how* it works. We also urge researchers to broaden the suite of outcomes typically measured and to explore new ways of empirically measuring behavioral change more directly. Finally, we encourage researchers to publish null findings. Without these findings, isolating what works in EE will continue to be elusive.

Acknowledgements

This work was supported by a grant from the National Education Council of the United States National Park Service.

Notes on contributors

Marc J. Stern is an associate professor in the Department of Forest Resources and Environmental Conservation where he teaches courses in environmental education and interpretation, social science research methods, and the human dimensions of natural resource management. His research focuses on human behavior within the contexts of natural resource planning and management, protected areas, and environmental education and interpretation.

Robert B. Powell is an associate professor in the Department of Parks, Recreation, and Tourism Management and the School of Agricultural, Forest, and Environmental Sciences. He teaches courses in recreation and protected areas management. His research focuses on education and communication, protected areas management, and sustainable tourism development.

Dawn Hill is an adjunct professor at the University of Arizona, where she teaches various psychology courses, two of which include environmental and conservation psychology. Additionally, through her nonprofit Ecologics, she performs summative and formative evaluations on environmental education programs. Her research focuses on pro/anti-environmental behavior and especially contextual influences that promote or impede that behavior.

References

- Agelidou, E., G. Balafoutas, and E. Flogaitis. 2000. "Schematisation of Concepts. A Teaching Strategy for Environmental Education Implementation in a Water Module Third Grade Students in Junior High School (Gymnasium – 15 Years Old)." *Environmental Education Research* 6 (3): 223–243.
- Aivazidis, C., M. Lazaridou, and G. F. Hellden. 2006. "A Comparison between a Traditional and an Online Environmental Educational Program." *Journal of Environmental Education* 37 (4): 45–54.
- Ajiboye, J. O., and S. A. Olatundun. 2010. "Impact of Some Environmental Education Outdoor Activities on Nigerian Primary School Pupils' Educational Knowledge." *Applied Environmental Education and Communication* 9 (3): 149–158.
- Ajzen, I. 2001. "Nature and Operation of Attitudes." *Annual Review of Psychology* 52: 27–58.
- Andrews, K. E., K. D. Tressler, and J. J. Mintzes. 2008. "Assessing Environmental Understanding: An Application of the Concept Mapping Strategy." *Environmental Education Research* 14 (5): 519–536.
- Ballantyne, R., J. Fien, and J. Packer. 2000. "Program Effectiveness in Facilitating Intergenerational Influence in Environmental Education: Lessons from the Field." *Journal of Environmental Education* 32 (4): 8–15.
- Ballantyne, R., J. Fien, and J. Packer. 2001. "School Environmental Education Programme Impacts upon Student and Family Learning: A Case Study Analysis." *Environmental Education Research* 7 (1): 23–37.
- Ballantyne, R., and J. Packer. 2002. "Nature-based Excursions: School students' Perceptions of Learning in Natural Environments." *International Research in Geographical and Environmental Education* 11 (3): 218–236.
- Ballantyne, R., and J. Packer. 2009. "Introducing a Fifth Pedagogy: Experience-based Strategies for Facilitating Learning in Natural Environments." *Environmental Education Research* 15 (2): 243–262.
- Basile, C. G. 2000. "Environmental Education as a Catalyst for Transfer of Learning in Young Children." *Journal of Environmental Education* 32 (1): 21–27.

- Baumgartner, E., and C. J. Zabin. 2008. "A Case Study of Project-based Instruction in the Ninth Grade: A Semester-long Study of Intertidal Biodiversity." *Environmental Education Research* 14 (2): 97–114.
- Bexell, S. M., O. S. Jarrett, X. Ping, and F. R. Xi. 2009. "Fostering Humane Attitudes toward Animals." *Encounter: Education for Meaning and Social Justice* 22 (4): 1–3.
- Bodzin, A. M. 2008. "Integrating Instructional Technologies in a Local Watershed Investigation with Urban Elementary Learners." *Journal of Environmental Education* 39 (2): 47–57.
- Bogner, F. X. 1999. "Empirical Evaluation of an Educational Conservation Programme Introduced in Swiss Secondary Schools." *International Journal of Science Education* 21 (11): 1169–1185.
- Bradley, J. C., T. M. Waliczek, and J. M. Zajicek. 1999. "Relationship Between Environmental Knowledge and Environmental Attitude of High School Students." *Journal of Environmental Education* 30 (3): 17–21.
- Braun, M., R. Buyer, and C. Randler. 2010. "Cognitive and Emotional Evaluation of Two Educational Outdoor Programs Dealing with Non-native Bird Species." *International Journal of Environmental and Science Education* 5 (2): 151–168.
- Cachelin, A., K. Paisley, and A. Blanchard. 2009. "Using the Significant Life Experience Framework to Inform Program Evaluation: The Nature Conservancy's Wings and Water Wetlands Education Program." *Journal of Environmental Education* 40 (2): 2–14.
- Carleton-Hug, A., and J. W. Hug. 2010. "Challenges and Opportunities for Evaluating Environmental Education Programs." *Evaluation and Program Planning* 33 (2): 159–164.
- Carrier, S. J. 2009. "Environmental Education in the Schoolyard: Learning Styles and Gender." *Journal of Environmental Education* 40 (3): 2–12.
- Culen, G. R., and T. L. Volk. 2000. "Effects of an Extended Case Study on Environmental Behavior and Associated Variables in Seventh- and Eighth-grade Students." *Journal of Environmental Education* 31 (2): 9–15.
- Cummins, S., and G. Snively. 2000. "The Effect of Instruction on Children's Knowledge of Marine Ecology, Attitudes toward the Ocean, and Stances toward Marine Resource Issues." *Canadian Journal of Environmental Education* 5: 305–326.
- D'Agostino, J. V., K. L. Schwartz, A. D. Cimetta, and M. E. Welsh. 2007. "Using a Partitioned Treatment Design to Examine the Effect of Project WET." *Journal of Environmental Education* 38 (4): 43–50.
- Dettmann-Easler, D., and J. L. Pease. 1999. "Evaluating the Effectiveness of Residential Environmental Education Programs in Fostering Positive Attitudes toward Wildlife." *Journal of Environmental Education* 31 (1): 33–39.
- DiEnno, C. M., and S. C. Hilton. 2005. "High School Students' Knowledge, Attitudes, and Levels of Enjoyment of an Environmental Education Unit on Nonnative Plants." *Journal of Environmental Education* 37 (1): 13–25.
- Dimopoulos, D., S. Paraskevopoulos, and J. D. Pantis. 2008. "The Cognitive and Attitudinal Effects of a Conservation Educational Module on Elementary School Students." *Journal of Environmental Education* 39 (3): 47–61.
- Eagles, P. F. J., and R. Demare. 1999. "Factors Influencing Children's Environmental Attitudes." *Journal of Environmental Education* 30 (4): 33–37.
- EECO (Environmental Education Council of Ohio). 2012. "Environmental Education Certification Competencies." <http://www.eeco-online.org/>
- Emmons, K. M. 1997. "Perceptions of the Environment While Exploring the Outdoors: A Case Study in Belize." *Environmental Education Research* 3 (3): 327–344.
- Ernst, J. 2005. "A Formative Evaluation of the Prairie Science Class." *Journal of Interpretation Research* 10 (1): 9–29.
- Ernst, J., and M. Monroe. 2004. "The Effects of Environment-based Education on Students' Critical Thinking Skills and Disposition toward Critical Thinking." *Environmental Education Research* 10 (4): 507–522.
- Farmer, J., D. Knapp, and G. M. Benton. 2007. "An Elementary School Environmental Education Field Trip: Long-term Effects on Ecological and Environmental Knowledge and Attitude Development." *Journal of Environmental Education* 38 (3): 33–42.

- Finn, A. N., P. Schrodt, P. L. Witt, N. Elledge, K. A. Jernberg, and L. M. Larson. 2009. "A Meta-analytical Review of Teacher Credibility and Its Association with Teacher Behaviors and Student Outcomes." *Communication Education* 58 (4): 516–537.
- Flowers, A. B. 2010. "Blazing an Evaluation Pathway: Lesson Learned from Applying Utilization-focused Evaluation to a Conservation Education Program." *Evaluation and Program Planning* 33 (2): 165–171.
- Gambino, A., J. Davis, and N. Rowntree. 2009. "Young Children Learning for the Environment: Researching a Forest Adventure." *Australian Journal of Environmental Education* 25: 83–94.
- Gastreich, K. R. 2002. "Student Perceptions of Culture and Environment in an International Context: A Case Study of Educational Camps in Costa Rica." *Canadian Journal of Environmental Education* 7 (1): 167–176.
- Grodzinska-Jurczak, M., A. Bartosiewicz, A. Twardowska, and R. Ballantyne. 2003. "Evaluating the Impact of a School Waste Education Programme upon Students', Parents' and Teachers' Environmental Knowledge, Attitudes and Behaviour." *International Research in Geographical and Environmental Education* 12 (2): 106–122.
- Ham, S. 2013. *Interpretation: Making a Difference on Purpose*. Golden, CO: Fulcrum.
- Hansel, T., S. Phimmavong, K. Phengsopha, C. Phompila, and K. Homduangpachan. 2010. "Developing and Implementing a Mobile Conservation Education Unit for Rural Primary School Children in Lao PDR." *Applied Environmental Education and Communication* 9 (2): 96–103.
- Heimlich, J. E. 2010. "Environmental Education Evaluation: Reinterpreting Education as a Strategy for Meeting Mission." *Evaluation and Program Planning* 33 (2): 180–185.
- Heimlich, J. E., and N. M. Ardoin. 2008. "Understanding Behavior to Understand Behavior Change: A Literature Review." *Environmental Education Research* 14 (3): 215–237.
- Hines, J. M., H. R. Hungerford, and A. N. Tomera. 1987. "Analysis and Synthesis on Responsible Environmental Behavior: A Meta-analysis." *Journal of Environmental Education* 18 (2): 1–8.
- Hollweg, K. S., J. R. Taylor, R. W. Bybee, T. J. Marcinkowski, W. C. McBeth, and P. Zoido. 2011. *Developing a Framework for Assessing Environmental Literacy*. Washington, DC: North American Association for Environmental Education.
- Hungerford, H., H. Bluhm, T. Volk, and J. Ramsey. 2001. *Essential Readings in Environmental Education*. 2nd ed. Champaign, IL: Stipes.
- Hungerford, H. R., and T. L. Volk. 1990. "Changing Learner Behavior Through Environmental Education." *Journal of Environmental Education* 21 (3): 8–21.
- Hungerford, H. R., T. Volk, and J. Ramsey. 2000. "Instructional Impacts of Environmental Education on Citizenship Behavior and Academic Achievement." Paper Presented at the 29th Annual Conference of the North American Association for Environmental Education, South Padre Island, TX, October 17–21. <http://www.cisde.org/pages/researchfindingspage/researchpdfs/IEEIA%20-%202020%20Years%20of%20Research.pdf>
- Hungerford, H. R., T. Volk, J. M. Ramsey, R. A. Litherland, and R. B. Peyton. 2003. *Investigating and Evaluating Environmental Issues and Actions*. Champaign, IL: Stipes Publishing, LLC.
- Iozzi, L. A. 1989. "What Research Says to the Educator. Part One: Environmental Education and the Affective Domain." *Journal of Environmental Education* 20 (3): 3–9.
- Jacobs, W. J., M. Sisco, D. Hill, F. Malter, and A. J. Figueredo. 2012. "On the Practice of Theory-based Evaluation: Information, Norms, and Adherence." *Evaluation and Program Planning* 35 (3): 354–369.
- Jacobson, S., M. D. McDuff, and M. C. Monroe. 2006. *Conservation Education and Outreach Techniques*. Oxford: Oxford University Press.
- Johnson, B., and C. C. Manoli. 2008. "Using Bogner and Wiseman's Model of Ecological Values to Measure the Impact of an Earth Education Programme on Children's Environmental Perceptions." *Environmental Education Research* 14 (2): 115–127.
- Johnson-Pynn, J. S., and L. R. Johnson. 2005. "Successes and Challenges in East African Conservation Education." *Journal of Environmental Education* 36 (2): 25–39.
- Klosterman, M. L., and T. D. Sadler. 2010. "Multi-level Assessment of Scientific Content Knowledge Gains Associated with Socioscientific Issues-based Instruction." *International Journal of Science Education* 32 (8): 1017–1043.

- Knapp, D., and G. M. Benton. 2006. "Episodic and Semantic Memories of a Residential Environmental Education Program." *Environmental Education Research* 12 (2): 165–177.
- Knapp, D., and R. Poff. 2001. "A Qualitative Analysis of the Immediate and Short-term Impact of an Environmental Interpretive Program." *Environmental Education Research* 7 (1): 55–65.
- Kruse, C. K., and J. A. Card. 2004. "Effects of a Conservation Education Camp Program on Campers' Self-reported Knowledge, Attitude, and Behavior." *Journal of Environmental Education* 35 (4): 33–45.
- Kuhar, C. W., T. L. Bettinger, K. Lehnhardt, S. Townsend, and D. Cox. 2007. "Evaluating the Impact of a Conservation Education Program." *IZE Journal of Natural Resources* 43: 12–15.
- Kusmawan, U., J. M. O'Toole, R. Reynolds, and S. Bourke. 2009. "Beliefs, Attitudes, Intentions, and Locality: The Impact of Different Teaching Methods on the Ecological Affinity of Indonesian Secondary School Students." *International Research in Geographical and Environmental Education* 18 (3): 157–169.
- Leeming, F. C., W. O. Dwyer, B. E. Porter, and M. K. Cobern. 1993. "Outcome Research in Environmental Education: A Critical Review." *Journal of Environmental Education* 24 (4): 8–21.
- Lindemann-Matthies, P. 2002. "The Influence of an Educational Program on Children's Perception of Biodiversity." *Journal of Environmental Education* 33 (2): 22–31.
- Liu, S., and M. S. Kaplan. 2006. "An Intergenerational Approach for Enriching Children's Environmental Attitudes and Knowledge." *Applied Environmental Education and Communication* 5 (1): 9–20.
- Louv, R. 2005. *Last Child in the Woods: Saving Our Children from Nature-deficit Disorder*. New York: Workman.
- Malandrakis, G. N. 2006. "Learning Pathways in Environmental Science Education: The Case of Hazardous Household Items." *International Journal of Science Education* 28 (14): 1627–1645.
- Martin, B., A. Bright, P. Cafaro, R. Mittelstaedt, and B. Bruyere. 2009. "Assessing the Development of Environmental Virtue in 7th and 8th Grade Students in an Expeditionary Learning Outward Bound School." *Journal of Experiential Education* 31 (3): 341–358.
- Mayer-Smith, J., O. Bartosh, and L. Peterat. 2009. "Cultivating and Reflecting on Intergenerational Environmental Education on the Farm." *Canadian Journal of Environmental Education* 14: 107–121.
- Middlestadt, S., M. Grieser, O. Hernández, K. Tubaishat, J. Sanchack, B. Southwell, and R. Schwartz. 2001. "Turning Minds on and Faucets Off: Water Conservation Education in Jordanian Schools." *Journal of Environmental Education* 32 (2): 37–45.
- Miller, D. L. 2007. "The Seeds of Learning: Young Children Develop Important Skills through Their Gardening Activities at a Midwestern Early Education Program." *Applied Environmental Education and Communication* 6 (1): 49–66.
- Monroe, M. C. 2010. "Challenges for Environmental Education Evaluation." *Evaluation and Program Planning* 33 (2): 194–196.
- Morgan, S. C., S. L. Hamilton, M. L. Bentley, and S. Myrie. 2009. "Environmental Education in Botanic Gardens: Exploring Brooklyn Botanic Garden's Project Green Reach." *Journal of Environmental Education* 40 (4): 35–52.
- NAAEE (North American Association for Environmental Education). 2012a. *Guidelines for Excellence*. Washington, DC: North American Association for Environmental Education. <http://eelinked.naaee.net/n/guidelines/posts/Download-Your-Copy-of-the-Guidelines>
- NAAEE (North American Association for Environmental Education). 2012b. *The Latest Research*. . Accessed November 1, 2012. <http://eelinked.naaee.net/n/eeeresearch>
- Nicolaou, C. T., K. Korfiatis, M. Evagorou, and C. Constantinou. 2009. "Developing of Decision-making Skills and Environmental Concern through Computer-based, Scaffolded Learning Activities." *Environmental Education Research* 15 (1): 39–54.
- Patton, M. Q. 2008. *Utilization-focused Evaluation*. 4th ed. Thousand Oaks, CA: Sage.
- PEEC (Place-based Education Evaluation Cooperative). 2012. *Place-based Ed. Research*. Accessed November 1, 2012. http://www.peecworks.org/PEEC/PEEC_Research/
- Poudel, D. D., L. M. Vincent, C. Anzalone, J. Huner, D. Wollard, T. Clement, A. DeRamus, and G. Blakewood. 2005. "Hands-on Activities and Challenge Tests in Agricultural and Environmental Education." *Journal of Environmental Education* 36 (4): 10–22.

- Powell, R. B., M. J. Stern, and N. Ardoin. 2006. "A Sustainable Evaluation Program Framework and Its Application." *Applied Environmental Education and Communication* 5 (4): 231–241.
- Powell, K., and M. Wells. 2002. "The Effectiveness of Three Experiential Teaching Approaches on Student Science Learning in Fifth-grade Public School Classrooms." *Journal of Environmental Education* 33 (2): 33–38.
- Powers, A. L. 2004. "Evaluation of One- and Two-day Forestry Field Programs for Elementary School Children." *Applied Environmental Education and Communication* 3 (1): 39–46.
- Randler, C., A. Ilg, and J. Kern. 2005. "Cognitive and Emotional Evaluation of an Amphibian Conservation Program for Elementary School Students." *Journal of Environmental Education* 37 (1): 43–52.
- Rickinson, M. 2001. "Learners and Learning in Environmental Education: A Critical Review of the Evidence." *Environmental Education Research* 7 (3): 207–320.
- Ruiz-Mallen, I., L. Barraza, B. Bodenhom, and V. Reyes-García. 2009. "Evaluating the Impact of an Environmental Education Programme: An Empirical Study in Mexico." *Environmental Education Research* 15 (3): 371–387.
- Russell, C. L. 2000. "A Report on an Ontario Secondary School Integrated Environmental Studies Program." *Canadian Journal of Environmental Education* 5: 287–304.
- Salata, T. L., and D. M. Ostergren. 2010. "Evaluating Forestry Camps with National Standards in Environmental Education: A Case Study of the Junior Forester Academy, Northern Arizona University." *Applied Environmental Education and Communication* 9 (1): 50–57.
- Schneider, B., and N. Cheslock. 2003. *Measuring Results*. San Francisco, CA: Coevolution Institute.
- Schneller, A. J. 2008. "Environmental Service Learning: Outcomes of Innovative Pedagogy in Baja California Sur." *Mexico. Environmental Education Research* 14 (3): 291–307.
- Siemer, W. F., and B. A. Knuth. 2001. "Effects of Fishing Education Programs on Antecedents of Responsible Environmental Behavior." *Journal of Environmental Education* 32 (4): 23–29.
- Sivek, D. J. 2002. "Environmental Sensitivity among Wisconsin High School Students." *Environmental Education Research* 8 (2): 155–170.
- Skibins, J. C., R. B. Powell, and M. J. Stern. 2012. "Exploring Empirical Support for Interpretation's Best Practices." *Journal of Interpretation Research* 17 (1): 25–44.
- Smith-Sebasto, N. J., and L. Cavern. 2006. "Effects of Pre- and Posttrip Activities Associated with a Residential Environmental Education Experience on Students' Attitudes toward the Environment." *Journal of Environmental Education* 37 (4): 3–17.
- Smith-Sebasto, N. J., and H. J. Semrau. 2004. "Evaluation of the Environmental Education Program at the New Jersey School of Conservation." *Journal of Environmental Education* 36 (1): 3–18.
- Sobel, D. 2012. "Look, Don't Touch: The Problem with Environmental Education." *Orion*, July/August.
- Stern, M.J., and R.B. Powell. (in press). "What Leads to Better Visitor Outcomes in Live Interpretation?" *Journal of Interpretation Research* 18 (2).
- 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." *Journal of Environmental Education* 39 (4): 31–43.
- Stern, M. J., R. B. Powell, and N. M. Ardoin. 2010. "Evaluating a Constructivist and Culturally Responsive Approach to Environmental Education for Diverse Audiences." *Journal of Environmental Education* 42 (2): 109–122.
- Tilden, F. 1957. *Interpreting Our Heritage*. Chapel Hill: University of North Carolina Press.
- UNESCO 1978. *Final Report Intergovernmental Conference on Environmental Education*, Tbilisi, USSR, 14–26 October 1977. Paris: UNESCO.
- Vaughan, C., J. Gack, H. Solorazano, and R. Ray. 2003. "The Effect of Environmental Education on Schoolchildren, Their Parents, and Community Members: A Study of Intergenerational and Intercommunity Learning." *Journal of Environmental Education* 34 (3): 12–21.

- Villegas, J. C., C. T. Morrison, K. L. Gerst, C. R. Beal, J. E. Espeleta, and M. Adamson. 2010. "Impact of an Ecohydrology Classroom Activity on Middle School Students' Understanding of Evapotranspiration." *Journal of Natural Resources and Life Sciences Education* 39: 150–156.
- Volk, T. L., and M. J. Cheak. 2003. "The Effects of an Environmental Education Program on Students, Parents, and Community." *Journal of Environmental Education* 34 (4): 12–25.
- Wright, J. M. 2008. "The Comparative Effects of Constructivist Versus Traditional Teaching Methods on the Environmental Literacy of Postsecondary Nonscience Majors." *Bulletin of Science, Technology, and Society* 28 (4): 324–227.
- Yager, R. E. 1991. "The Constructivist Learning Model: Towards Real Reform in Science Education." *The Science Teacher* 9: 53–57.
- Zelezny, L. C. 1999. "Education Interventions That Improve Environmental Behaviors: A Meta-Analysis." *Journal of Environmental Education* 31 (1): 5–14.
- Zint, M. 2012. *My Environmental Education Evaluation Resource Assistant*. Accessed November 1, 2012. <http://meera.snre.umich.edu/>
- Zint, M., A. Kraemer, H. Northway, and M. Lim. 2002. "Evaluation of the Chesapeake Bay Foundation's Conservation Education Programs." *Conservation Biology* 16 (3): 641–649.

Appendix

Table A. Coincidence of program characteristics with any positive findings for each measured outcome.

Reported practices	Desired Outcomes							Total
	Knowledge	Skills	Attitudes	Intention	Behavior	Enjoyment	Total	
Imm	14	9	7	3	5	4	42	
% Positive Finding	100%	100%	100%	100%	100%	100%	100%	
PI	3	3	2	1	3	1	13	
% Positive Finding	100%	100%	100%	100%	100%	100%	100%	
DC	25	8	15	5	9	7	69	
% Positive Finding	100%	100%	93%	100%	100%	100%	99%	
Proj	19	12	11	6	6	6	60	
% Positive Finding	100%	92%	100%	83%	100%	100%	97%	
GI	17	3	9	1	5	2	37	
% Positive Finding	100%	100%	89%	100%	100%	100%	97%	
Inv	35	15	23	10	11	12	106	
% Positive Finding	100%	100%	87%	90%	100%	100%	96%	
Ref	29	11	19	6	9	7	81	
% Positive Finding	100%	100%	84%	83%	100%	86%	94%	
View	6	3	6	4	5	5	29	
% Positive Finding	100%	100%	83%	75%	100%	100%	93%	
HO	52	20	32	16	16	12	148	
% Positive Finding	98%	95%	81%	81%	100%	100%	93%	
Issue	36	17	26	14	12	11	116	
% Positive Finding	100%	88%	85%	79%	100%	91%	91%	
Outside	48	14	43	17	15	19	156	
% Positive Finding	98%	100%	81%	82%	93%	95%	91%	
Place	43	18	36	17	15	15	144	
% Positive Finding	98%	94%	83%	82%	93%	93%	91%	
C/G	28	10	19	7	10	10	84	
% Positive Finding	100%	100%	74%	71%	100%	90%	90%	

(Continued)

Table A. (Continued).

Reported practices	Desired Outcomes							Total
	Knowledge	Skills	Attitudes	Intention	Behavior	Enjoyment		
AP	Observations	63	24	39	18	17	14	175
	% Positive Finding	95%	92%	82%	72%	94%	100%	90%
Rel	Observations	18	7	16	6	8	8	63
	% Positive Finding	100%	86%	75%	83%	100%	88%	89%
LC	Observations	7	5	6	2	3	4	27
	% Positive Finding	100%	100%	83%	50%	100%	75%	89%
Mod	Observations	53	16	37	17	12	15	150
	% Positive Finding	94%	94%	76%	71%	92%	93%	87%
Play	Observations	12	3	8	5	4	5	37
	% Positive Finding	83%	100%	63%	60%	75%	100%	78%
Lect	Observations	24	6	17	3	7	6	63
	% Positive Finding	88%	100%	47%	33%	86%	83%	75%
Inside	Observations	21	9	10	6	2	1	49
	% Positive Finding	86%	78%	40%	50%	50%	0%	67%
Trad	Observations	5	0	5	1	1	2	14
	% Positive Finding	40%	NA	0%	0%	0%	50%	21%
Total	Observations	558	213	386	165	172	162	1656
	% Positive Finding	96%	95%	79%	77%	95%	93%	90%

Notes: If a particular outcome was mixed, then it was coded as 'positive' here. For example, if three out of four attitudes showed positive results for a program, then that program's characteristics would be coded to show a 'positive finding' in this table.