Reconceptualizing 'effectiveness' in environmental projects: Can we measure values-related achievements?

CrossMark

Marie K. Harder [a,b,*], Ismael Velasco [b,1], Gemma Burford [b,1], Dimity Podger [b,2], Svatava Janoušková [c,3], Georgia Piggot [b], Elona Hoover [b,1]

[a] Department of Environmental Science and Engineering, Fudan University, 220 Handan Road, Shanghai 200433, China
[b] Sustainable Development Coordination Unit, Faculty of Science and Engineering, University of Brighton, Cockcroft Building, Lewes Road, Brighton BN2 4GJ, UK
[c] Charles University Environment Center, José Martího 2/407, 16200 Prague 6, Czech Republic

ARTICLE INFO

ABSTRACT

There have been recent calls for a shift to an evidence-based paradigm in environmental management, grounded in systematic monitoring and evaluation, but achieving this will be complex and difficult. Evaluating the educational components of environmental initiatives presents particular challenges, because these programs often have multiple concurrent goals and may value 'human outcomes', such as value change, which are intangible and difficult to quantify. This paper describes a fresh approach based on co-creating an entirely new values-based assessment framework with expert practitioners worldwide. We first discuss the development of a generic framework of 'Proto-Indicators' (reference criteria constituting prototypes for measurable indicators), and then demonstrate its application within a reforestation project in Mexico where indicators and assessment tools were localized to enhance context-relevance. Rigorously derived using unitary validity, with an emphasis on relevance, practicability and logical consistency from user perspectives, this framework represents a step-wise advance in the evaluation of non-formal EE/ESD programs. This article also highlights three important principles with broader implications for evaluation, valuation and assessment processes within environmental management: namely peer-elicitation, localizability, and an explicit focus on ethical values. We discuss these principles in relation to the development of sustainability indicators at local and global levels, especially in relation to post-2015 Sustainable Development Goals.

1. Introduction

The need for an 'effectiveness revolution' in environmental management, inspired by transitions to evidence-based management in fields such as medicine and public health, is increasingly being recognised (Ferraro and Pattanayak, 2006; Heimlich, 2010; Keene and Pullin, 2011; Rode and Michelsen, 2008; Sutherland et al., 2004). Governments and donors alike are waking up to the fact that, on the whole, "the organizations to which the public pays and donates billions of dollars cannot yet demonstrate their effectiveness at providing future generations with a healthy environment" (Keene and Pullin, 2011, p. 2134). The absence of clearly articulated program objectives and evaluation criteria is a pervasive problem in non-formal environmental education (EE) and 'education for sustainable development' (ESD) initiatives, whether conducted in isolation or within broader natural resource management programs, (Carleton-Hug and Hug, 2010; Heimlich, 2010), and also extends beyond environmental contexts to other types of non-formal education (Christensen et al., 2005; Wiltz, 2005).

In order for the revolution to succeed, it is necessary to identify *how* the transition to an evidence-based paradigm of environmental management might be accomplished. As highlighted by Springett (2001, 2003), two fundamental questions in any evaluation activity are, first, whose values are driving the evaluation, and second, against whose standards the project activities are measured. More broadly, one might ask whose values underpin the

* Corresponding authors. Department of Environmental Science and Engineering, Fudan University, 220 Handan Road, Shanghai 200433, China. Tel.: +44 1273 644707.
E-mail addresses: M.K.Harder@brighton.ac.uk (M.K. Harder), ismael@adorafoundation.org (I. Velasco), G.L.Burford@brighton.ac.uk (G. Burford), dimity@barasaconsult.com (D. Podger), svatava.janouskova@czp.cuni.cz (S. Janoušková), georgiapiggot@gmail.com (G. Piggot), e.hoover@brighton.ac.uk (E. Hoover).
[1] Tel.: +44 1273 644707.
[2] Tel.: +61 2 9799 6007.
[3] Tel.: +420 251080344.

initiative or program that is being evaluated. If such questions are not systematically considered, evaluation criteria selected by more powerful stakeholders will often be adopted by default, while the worldviews and priorities of the marginalized may be neglected.

There is much to be learnt from current trends in evaluation studies, such as the development of process-based, participatory, utilization-focused, empowerment-oriented and collaborative evaluation approaches (Burford et al., 2013b; Crishna, 2007; Daigneault and Jacob, 2009; Donaldson et al., 2010; Ellis and Hogard, 2006; Fitzpatrick, 2012; Flowers, 2010; Hogard, 2008; Holte-McKenzie et al., 2006; Springett, 2003). A positive precedent has also been set within the arena of assessing land degradation. Here, new adaptive learning processes for indicator development bring together 'top-down' (expert-led) and 'bottom-up' (community-based) approaches, with a view to achieving a balance between objectivity and ease-of-use criteria (Fraser et al., 2006; Reed et al., 2005, 2006; Stringer and Dougill, 2013). In this paper, we have applied the adaptive learning principle specifically to the evaluation of non-formal environmental education, with a focus on programs whose objectives are values-based rather than purely biophysical, and developed a values-based framework with the potential for broader application.

The work reported here was initiated by demand from civil society organizations (CSOs) who wanted to find context-relevant, practicable and local ways to assess their educational work in sustainable development. For them, assessment of learning could not be separated from overall program evaluation because of the breadth of learning taking place. Rather than imposing external frameworks for evaluation and learning assessment, built on different premises, we decided to build from scratch a new, peer-elicited framework designed to validly represent the worldviews of these practitioners. This was achieved through a consortium approach in which CSO representatives and academic researchers worked together as equal partners, with CSOs holding the balance of decision-making power (Podger et al., 2010).

In Section 2 we present the need for, and challenges of, evaluation in non-formal environmental education (EE) and 'education for sustainable development' or 'education as sustainable development' (ESD). This necessitates unitary validity guidelines and a values-lens approach, so we present background sections on these (Sections 2.2.1 and 2.2.2 respectively). The peer-elicitation of the new framework is then outlined in Section 3.2, and its practical application is described in Section 3.3. A discussion is then provided to relate the results to the initial questions, to highlight significant contributions to other fields of work, and finally (in Section 5.1) to reflect briefly on broader implications for environmental management and global sustainability assessment.

2. Background

2.1. The need for appropriate evaluation of non-formal EE/ESD

For the purpose of this paper, we will use the European Centre for Vocational Training (Cedefop, 2001) definition of *non-formal education* as planned activities which contain an important learning element and are intentional from the learner's point of view, but may not necessarily be explicitly designated as learning activities.[4] In EE/ESD contexts, non-formal education may include community-based activities with an explicit or implicit learning element (such as reforestation, local habitat management or wildlife survey projects), as well as more structured initiatives that primarily target young people outside school hours (e.g. 'Forest Schools', 'Wildlife Clubs' or summer camps).

Mainstream formal education has decades of co-evolution between learning assessment methods and underlying theories of learning and knowledge, providing a firm foundation for evaluation structures. Non-formal education, however, currently lacks any such infrastructure (Clavijo et al., 2005). This gap has become more problematic in recent years, as the societal roles played by non-formal education worldwide have increased in importance and scope (Carleton-Hug and Hug, 2010; Keene and Pullin, 2011). Additionally, funding bodies and donors are increasingly demanding measurable accountability, and becoming frustrated by the inadequacies of conventional evaluation approaches (e.g. Ford Foundation, 2011; Wightman, 2010).

This need is especially pressing and well-documented in the case of non-formal EE and ESD initiatives. These are considered close cousins to formal education, often with learning objectives that appear similar, yet EE is reportedly struggling with rigorous evaluation, as described in a special issue of *Evaluation and Program Planning* (see Crohn and Birnbaum, 2010 for overview). Most EE programs do not incorporate evaluation into their activities (Fleming and Easton, 2010), and practitioners are often unaware of applicable literature to support their goals and methods (Wiltz, 2005). The research base is relatively undeveloped, relying largely on related formal disciplines (Wiltz, 2005) but 'borrowing' concepts from formal education is not necessarily appropriate, because fundamental learning objectives may differ even when the content appears similar.

Non-formal EE/ESD programs present many difficulties to evaluators. They often have multiple goals (Christensen et al., 2005) — some not focused on content — and the individual learner is not always the most appropriate level of measurement. For example, the intended beneficiary of an educational intervention may be a specific ecosystem, or 'nature' in general: the crucial question may not be how much an individual knows about water conservation, but how much water is conserved within a community (Heimlich, 2010). Furthermore, 'situated learning' within a community of practice demands 'situated assessment', e.g. assessing group members' ability to consult with one another and work together to solve problems (Singh, 2011). Many of the goals of non-formal EE and ESD are fundamentally difficult to translate into measurable outcomes: they may be long-term, broad and poorly defined, and, crucially, *affective* in nature. Furthermore, as Wiltz (2005: 18) explains: "The alternative approaches and settings of non-formal education are intended to foster often very personal outcomes in each of the participants." Assessors sometimes refer to these as 'unintended' outcomes, but organizers may view them as fundamental.

Evaluation of non-formal programs can involve a larger, more varied and less defined set of variables, such as changes in participant relationships, levels of participation, feelings of belonging (Christensen et al., 2005) — each requiring different assessment constructs and methods. Furthermore, each non-formal program emphasizes different sets of results, making comparisons across projects very difficult.

Altogether, these challenges provide huge barriers to non-formal EE and ESD evaluation because any proposed framework must be complex and multi-faceted to cope with the range of variables. Additionally, such a framework cannot directly rely on concepts of formal education assessment as they focus on individual content learning and thus are not appropriate for evaluation of non-formal education.

---

[4] Non-formal education differs from formal education in that the latter involves explicitly recognized learning activities which are conducted within formal institutions, such as schools or universities, and have predefined learning outcomes, resources and time. It is also distinct from informal learning, which is unintentional from the learner's perspective and occurs during daily work or leisure activities, e.g. via the media (Cedefop, 2001).

Civil society organizations (CSOs) have long expressed their frustration with conventional donor-driven approaches to evaluation, such as the Logical Framework Approach, which focus on easily quantifiable 'products' (e.g. the number of trees planted or workshops conducted) in preference to innovative process-driven approaches (Ebrahim, 2002, 2003; Edwards and Hulme, 1996; Riddel, 1999). Such a focus has often resulted in the imposition of bureaucratic monitoring and evaluation systems that prove to be more hindrance than help (Crishna, 2007), while also leaving many CSOs struggling to articulate − let alone evaluate − less tangible aspects of what they value. This issue has come to a head in recent years, with the growing demand for a new generation of indicators to reflect the distinctive educational mandate of ESD, characterized by interdisciplinarity, multiperspectivity and participation (Rode and Michelsen, 2008). The United Nations Economic Commission for Europe (UNECE), for example, has called for strategies and instruments to assess changes in knowledge, attitude and practice as a result of non-formal and informal learning among citizens in relation to ESD (UNECE, 2005, 2006).

There is thus a clear need, becoming acute in some areas, for effective evaluation systems for non-formal EE and ESD programs. Barriers to effective evaluation seem to arise at three levels: (a) the goals of non-formal education diverge from traditional teaching and learning, and do not have well-developed underpinning academic theories against which they can be directly evaluated; (b) appropriate indicators for the goals do not exist, and cannot be simply transferred from other fields; and (c) suitable assessment approaches for candidate indicators appear difficult to develop. Thus, although academic literature can offer *concepts* relevant to the evaluation of non-formal education, specifically applicable details are largely lacking. Significant progress can be made only by facilitating the participation of, and formalizing the knowledge of, organizations which deliver non-formal EE and ESD (here denoted as 'CSOs' for simplicity, although we acknowledge that some environmental management initiatives with educational components are led by public sector institutions).

There is, however, another barrier which is relevant not only to creating but also to applying evaluation frameworks: many non-formal programs do not have well-defined goals (Wiltz, 2005). It is understandable that goals may not be easily articulated in relation to formal learning objectives, but smaller CSOs in particular may be unable to specify their goals in *any* terms − although some still report confidently, on the basis of instinct, that their work was 'successful'. This signals that goal clarification is needed as part of any effective evaluation system for CSOs: otherwise, the entire basis of the evaluation may be invalid. This in turn implies, again, that deep participation of CSO stakeholders is likely to be a vital element in every evaluation.

In summary, literature and experience indicate the clear need for effective assessment systems in non-formal EE/ESD, which should have the following characteristics: (a) flexibility, to accommodate the great variability of parameters; (b) rigour, to produce valid results; (c) an emphasis on participation by CSO representatives; (d) meaningfulness, relevance and practical utility within CSO contexts; and (e) the ability to contribute to goal definition or clarification, where appropriate.

With these concerns in mind, researchers at the University of Brighton initiated exploratory discussions with local and international CSOs involved in non-formal EE/ESD. It became clear that the CSOs saw their work and goals as 'values-based' (as in 'ethical values', e.g. trust, respect, equality and empowerment). The 'learning' that interested them was defined in terms of integrating these values in society, rather than specific knowledge transfer

events. These ethical values are generally considered to be intangible and therefore unable to be weighed, measured or counted directly (Schlater and Sontag, 1994), with the result that CSOs have been evaluated with externally-imposed frameworks that often overlooked what they regarded as core goals and achievements. However, as first noted by Handy (1970), behaviours and practices associated with values can be observed and measured (see also Burford et al., 2012; Burford et al., 2013b; Podger et al., 2013). After consultations with a range of CSOs, a consortium of CSO representatives and academic researchers was formed to explore whether useful values-based indicators could, in principle, be developed. This paper reports on that work from the EU project ESDinds.[5]

The insistence of the CSOs that their work required a values lens, and the fact that they could not provide a ready-made framework of values and indicators, signaled that our research would need to focus on carefully identifying and defining these values in context. Three factors would be key: careful considerations of validity to ensure any results were rigorous; an understanding of values; and deep participation of the CSOs, as "experts on their experience domains", within the research co-design process (Sleeswijk Visser et al., 2005, 127; Steen et al., 2011). These issues are discussed in the sections that follow.

2.2. Background frameworks for the research

2.2.1. Validity

Our participating CSOs had indicated that their broad educational work, while focussing on environmental and/or social outcomes associated with sustainable development, was primarily values-based. But if values-based achievements are considered intangible (Schlater and Sontag, 1994), initial work is required to conceptualize them, and particular care must be taken to ensure that any measurements are clearly representative of them. This begs the question of whether the final interpretations of such 'measurements' can ever have any validity. Clearly, validity issues needed careful consideration in this work, and a summary of these is presented.

Formal education is built on conceptual frameworks where elements of knowledge, skills and competencies sit in known relationship to each other, with established links to underlying substantive theories (such as how children learn, or processes of cognition). Assessment in formal education is thus not only easily framed in its own context (important to 'content' validity, see below), but also with respect to underpinning theories (needed for 'substantive' validity, see below). For most types of non-formal education, by contrast, underlying theories are absent or undeveloped, and the variety and variability of types of learning involved appears to preclude any unitary frameworks (Wiltz, 2005). This leaves wide open the question of what can be assessed in non-formal education with any validity.

To address this issue, we carefully review general concepts of assessment validity, with this application in mind. One of the most influential authors in this field is Messick (1989), who describes validity as "an overall evaluative judgement of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of interpretations and action on the basis of…modes of assessment" (Messick, 1989). According to this

---

'unitary validity' approach, the concept of validity is not applicable to assessments or tools as such, but to the *meanings* given to assessment results, with the corollary that the end use of those results must also be considered (American Educational Research Association, 1999; Messick, 1995). In other words, validity must be checked on a case-by-case basis: a valid use of a tool's results for one purpose may be invalid for another. Every link in the chain, back to the related core constructs, must be checked for validity with respect to the *use* of the interpretation. Unitary validity has six closely interrelated aspects, as described in published validity standards (American Educational Research Association, 1999; see also Messick, 1995):

a) Content aspect — given that a specific domain (e.g. knowledge, or the enactment of ethical values) is being assessed, is the assessment within its boundaries? Is it sufficiently representative for the assessment question? Is there an appropriate framework that the assessed items are known to fit into, e.g. algebra into mathematics; honesty into a framework of other values?

b) Substantive aspect — how do the assessed items relate to underpinning substantive theories, e.g. process models of task performance? Are some processes being over-represented? For example, if testing for verbal aptitude, does the test design take into account competing impacts from analytical aptitude?

c) Structural aspect — is the weighting given to sub-results congruent with the theories involved, e.g. if IQ is being tested, are the component aptitudes weighted appropriately?

d) Generalizability aspect — can the results be generalized across other samples, e.g. with different settings or environments?

e) External aspect — are the assessment results consistent with results from separate external tests, e.g. based on constructs that should give divergent or convergent results?

f) Consequential aspect — does the assessment contain everything needed to ensure that the final use of the results is valid, i.e. that any decisions made on the basis of the assessment are appropriate? (For example, if the number of youth attending is used to determine success of a program, but the CSO's aim is developing fewer but deeper relationships to engender behaviour change, then consequential validity is in question.)

These different aspects of validity are only useful when considered as part of an overall unitary concept — the degree that the six aspects support the intended interpretation of the proposed assessment purpose. All relevant aspects of validity should be checked. Unfortunately, many published studies focus only on some aspects (e.g. statistical evidence for sampling), and fail to properly consider others (e.g. whether the items measured are sufficiently relevant to the question posed). This may result in an overall lack of validity.

The above framework for validity is well established and can be applied to measurements generally, whether in quantitative science, or e.g. psychology and education, where they originated (Messick, 1995). The six aspects can be reorganized into different categories and sometimes even different numbers of categories depending on the approach taken in a given discipline, but however they are organized, each set represents what is needed for overall, or 'unitary validity for the intended use. Terms such as *test score*, *assessment result* or *measurement* can be interchanged as appropriate; all refer to the thing which is measured. For an assessment tool to be capable of providing a valid result, it is vital that it does not inherently build in specific invalidating steps (e.g. a weakness in the rationale that designates one variable as an indicator of another), nor general steps which prevent validation checks later (e.g. using the test for a new purpose without checking the implications for consequential validity).

The aim of our work was to develop appropriate indicators that could be used by CSOs to monitor values-based achievements which constitute, or contribute to, the success of their endeavours. In this sense, we were developing indicators for the principles, practices, and outcomes that represent the operational expression of the CSOs' ethical values (c.f. Anello, 2006). This intimate link with the CSOs was the central guiding principle for designing our evaluation tool, and it was used to prioritize aspects of validity which are specific to our design, resulting in the following reference principles (which are applicable regardless of which values framework is eventually used):

I. Relevance to the CSO — the values (and proxies for them) must be directly relevant and important to the CSOs involved, and the framework they lie in should not be dissonant with the CSO's values (cf. content and structural aspects of validity).

II. Practicable — The measurements (assessments) to be made should be practical and feasible for the CSO, and sympathetic to their usual activities.

III. Valid links — The CSO should be satisfied with every link in the chain of rationale between the value, the indicator, the measurement (assessment) and its use.

Note the emphasis on consequential and content validity, which are framed with respect to the CSO point of view. Substantive, structural and external validities cannot be emphasized until underlying frameworks are specified, and in their absence would have to be judged via the CSO's qualitative interpretation. As the focus of the work at this stage is feasibility (i.e. using only a few CSOs) rather than transferability, then generalizability validity would only be relevant at a later stage, if initial work is successful. (See Section 3.2.4 for more on generalizability.)

In outlining Messick's (1995) six core aspects of validity, derived initially from the positivist paradigm, we are not implying any rejection of alternative, post-positivist conceptualizations of validity emerging from qualitative approaches to research (Lincoln and Guba, 1985, 1986). Rather, as Cohen et al. (2011, p. 183) emphasize, whilst there may be different canons of validity in different types of research, they are not mutually exclusive and there is common ground between them. This is evidenced by the emergence of mixed methods research, which begins with a strong mixed research question such as 'what and how' (Tashakkori and Creswell, 2007, p. 207) and seeks to integrate qualitative and quantitative components in such ways as to make them mutually illuminating (Bryman, 2007). Mixed methods research includes the principles of 'paradigmatic mixing legitimation' (the extent to which combining the ontologies, epistemologies and methodologies of different research paradigms can generate useful results); and 'political legitimation' (the extent to which meta-inferences stemming from the combination of methods are accepted by the audiences) (Onwuegbuzie et al., 2007). The focus on useful results and acceptance by audiences reflects the pragmatist paradigm in mixed methods research, premised on the understanding that the goal of research is to identify practical solutions to real-world problems (Onwuegbuzie and Leech, 2005).

In summary, a consideration of validity issues has shown the need for the ESDinds project to produce indicators with these qualities: seen by the CSOs (I) to be relevant; (II) to be practicable; and (III) to have valid logical links to their own core values.

We should note that there are situations where additional requirements might be needed to achieve valid evaluations, for example with the involvement of donors, or specific moves towards the 'effectiveness revolution'. In either case the purpose of the measures changes, requiring different ways forward. For example, from a large set of indicators meeting I, II and III above, the CSO might naturally only chose favourable ones for actual measurement: the donors might thus specify their own choices. The CSO might veer towards indicative measures of chosen indicators: donors or the public might require more rigorous measures including multiple assessment methods which achieve some triangulation. We discuss some possible related developments to achieve this in the Discussion.

2.2.2. Values

As our partner CSOs claimed that their projects and thus their results were linked to ethical values, it was important to determine whether any useful frameworks already existed for understanding and measuring these values (irrespective of whether they were directly linked to individual learning). This section gives an overview of the field, bearing in mind the project's principles I, II and III set out in Section 2.1 above.

There is no universally agreed theoretical concept of values, in spite of the attention paid by sociologists and social psychologists to values and value measurement over several decades. An attempt to summarize and clarify the concept in social science literature identified 180 different definitions of 'value' (Horáková, 2005). We note, however, that the definitions by Kluckhohn, Williams and Rokeach have been particularly influential. Kluckhohn (1951, p. 395) defined a value as "a conception, explicit or implicit, distinctive of an individual or characteristic of a group, of the desirable, which influences the selection from available modes, means and ends of action", while Williams (1979, p. 16) defined values as "criteria or standards of preference", noting that all values have cognitive, affective and directional aspects. Rokeach (1973, p. 5) defined the value concept as "an enduring belief that a specific mode of conduct or end-state of existence is personally or socially preferable to an opposite or converse mode of conduct or end-state of existence", albeit with an understanding that values may change over time and are part of a dynamic system with inherent contradictions.

The first assessment tool for 'measuring values' was the All-port–Vernon–Lindzey Study of Values (SOV) (Allport et al., 1951). The SOV was based on Spranger's (1928) six 'ideal types' of value orientations (theoretical, economic, political, aesthetic, social and religious), first published in German in 1914; it has had a lasting impact, and has recently been updated for the 21st century (Kopelman et al., 2003). Still more prominent is the Rokeach Values Survey (RVS), which operationalizes values by using a simple list of 18 instrumental values (referring to modes of conduct, e.g. *ambitious*, *clean*, *independent* and *loving*), and 18 terminal values (referring to idealized end-states of existence, e.g. *a comfortable life*, *equality*, *health*, and *mature love*). The RVS asked the respondent to rank the 18 value-labels[6] in each set in terms of personal importance (Rokeach, 1973, 1979). Researchers then relate the answers back to their pre-determined framework of the 36 values, within which each value has its own definition and clear boundaries with respect to the others. From this, the researchers make inferences about the values held by the respondents. In this measure, each value is represented by a single value-label, and the 36-item values

framework is set by the researchers and unknown to the respondents.

More recently, Schwartz (1994) defined values as desirable, *trans-situational* goals that people strive to attain (in contrast to 'norms' and 'attitudes', which refer to specific actions, objects, or situations). Building on the work of Rokeach, Schwartz and colleagues identify ten distinct value orientations, intended to include all core values recognized in cultures around the world: power, achievement, hedonism, stimulation, self-direction, universalism, benevolence, tradition, conformity and security (Sagiv and Schwartz, 1995; Schwartz, 1994). This concept is operationalized in the Schwartz Values Survey (SVS) which presents 57 single value-label items to respondents, who are asked to rate them as guiding principles in their lives by using a 9-point scale from 7 ('of supreme importance') to 0 ('of no importance') and −1 ('opposed to my values') (Schwartz, 1992, 1994). Again, each value is represented by a single value-label (word or phrase), and the values framework is unknown to the respondents.

The Portrait Values Questionnaire (PVQ) is a variation developed by Schwartz et al. (2001, p. 522) in response to concerns that certain populations might struggle with the abstract thinking demanded by the original SVS. In place of single value-labels, participants are asked to rate verbal 'portraits' of hypothetical individuals. For example, the portrait statement for the 'self-direction' value is: "Thinking up new ideas and being creative is important to him. He likes to do things in his own original way." The use of such statements negates the need for the respondent to interpret the simple value-label ('self-direction'); but on the other hand, the researchers have effectively imposed their own interpretation, in addition to using an externally set values framework not known to the respondents.

The models mentioned above can all be termed 'closed models' in that they offer respondents a closed set of items to respond to (Schlater and Sontag, 1994). The fixed nature of such tools allows comparability across applications − they can be used in a standardized manner for varied populations. They all make use of a rigid values framework pre-defined by the researchers, built on theoretical work originating in earlier decades. These models are not suited to our work with CSOs and non-formal EE/ESD because of their rigidity and lack of connection with CSO enacted values; they violate the requirement for validity to be case-specific, which in turn requires that the framework be known by and acceptable to the CSOs.

Schlater (1969) devised a different approach that used open questions to elicit individualized values statements from respondents after presenting projective stories about family decisions, by asking the question "what should be done, and why?". This open questioning allows the respondents to articulate values in their own terms, rather than being limited by a closed, prescribed framework. In this particular study, however, an external framework was still imposed *after* data collection by the researcher, who categorized the resulting statements into four broad value types, i.e. traditional, social, autonomous, and change-prone. There is every possibility that the respondents' interpretation of their own responses might diverge from those of the researcher, in contrast to the situation we are trying to achieve in this work, i.e. deep 'face validity' to the user CSOs. Thus, while the open method of data *collection* is useful for our purposes, the data analysis step employed by Schlater is potentially problematic.

A third model type, with an in-built feedback loop for respondents, was modified from the Competing Values Framework (CVF) (Cameron and Quinn, 1999). The CVF, which has been widely used to measure organizational values in the public sector (Parker and Bradley, 2000) and higher education (Silver, 2003), asks respondents to divide and allocate 100 'points' between four

---

[6] In this paper and elsewhere (e.g. Burford et al., 2013b), we use the term 'value-labels' as a synonym for 'words and phrases used to represent values', for the purpose of convenience.

scenarios representing different aspects of organizational culture — allocating the most points to the scenario that best resembles their own organization. The modified version, known as the Organizational Values Questionnaire (OVQ) (Reino and Vadi, 2010), has been developed in Estonia. In translating and adapting the original CVF to the Estonian context, 21 keywords in the Estonian language were selected to reflect the four value categories used in the CVF, and several questionnaire items were constructed for each keyword. Three potential users of the system (who were Estonian, taking the role of 'experts' of their own assessment needs) were asked to comment on the clarity of these questionnaire items, and to assign one or more items to each of the keywords as a backwards check. Items with good inter-rater reliability and no deviation from the model were submitted to a second panel of ten Estonians, who were asked to rate how well each draft questionnaire item fitted its related value type. This feedback was used to improve the items.

The OVQ is distinctive among values measurement tools for two reasons: its involvement of potential users of the system to 'localize' the indicators so that respondents would relate to them better, and its inclusion of a 'feedback loop' to validate value-indicator links. This concept of 'localization' is useful in our work, where it can be applied to a process of bridging the viewpoints of our own potential 'users' (the CSOs) and those of the researchers trying to formalize values concepts. Nonetheless, as the OVQ is still based on an underlying researcher-generated framework of four value types, it was not wholly appropriate for our purpose of developing assessments with optimal face validity.

In summary, current literature on the measurement of values indicates that mainstream models use frameworks that are externally set by researchers without user involvement; are often based on understandings of values and their structure that are many decades old; and usually involve closed questions, with interpretation by researchers. These characteristics may be appropriate for the original uses of those models, but they do not match our reference principles I, II and III for assessments to be relevant, practicable and validly linked to CSO values. These can only be met through deep involvement of CSOs in developing the values framework, the assessment methods and the interpretations. Thus, in co-designing our own approach to satisfy the three reference principles, we realized that our task was not merely to develop indicators: we would have to also develop a valid values framework to base them on ('Phase 1' of the project), and develop usable assessment tools ('Phase 2'). Some lessons gleaned from existing values assessment systems were useful in this co-design process, such as the use of open questions (c.f. Schlater, 1969) and localization via a feedback loop using appropriate 'experts' (c.f. Reino and Vadi, 2010); yet the need remained for us to develop an innovative approach, as discussed in Section 3, rather than working to an existing protocol.

3. Research paradigm and methodology

The research paradigm was informed by critical social theory and the epistemological approach found in Lather's (1986) seminal essay *Research as Praxis*. Three central principles of Lather's research as praxis are particularly relevant: (i) critical inquiry is a fundamentally mutually educative enterprise; (ii) it invites the critical reaction of participants; and (iii) it stimulates a self-sustaining process of critical analysis and enlightened action. These principles come from the critical epistemological assumption that knowledge is socially constituted, contextual, and value-laden (Lather, 1986) and recognition of the knowledge/practice dialectic; that is, knowledge is generated through practice, and practice is itself informed by knowledge. Further, knowledge both catalyses and is generated by changes in practice (Wainwright, 1997).

Through research-as-praxis, the critical researcher engages participants in dialogue in order to uncover underlying meanings and influences of their social practices, vision and goals. Together they develop an understanding of their cultural and social environment and associated values, i.e. practical knowledge production (Habermas, 1971). A reciprocal relationship, where each is both an educator and a learner, is necessary (Gramsci, 1971; Lather, 1986). This tenet is consistent with the approach needed in our research, where the researchers did not initially have a deep understanding of CSO practices, and the CSOs did not have a conceptualized framework for the values underpinning their work. Thus, both needed each other, in dialogue, to make progress in building a relevant, practical and valid evaluation framework for non-formal EE/ESD.

Lather (1986) also argues that research participants should have the opportunity to critically react to and test the validity of the researcher's conception and theoretical explanations about their world. This implies that they need to be involved in generating those interpretations, essential to develop such a new conception about a sphere of experience and practice that they are 'experts' of (see also Kushner and Norris, 1980; Steen et al., 2011).

3.1. Specific research approach

In line with a research-praxis, dialogic, interactive approach, the two-year ESDinds project was set out in three major iterations. It incorporated significant review, planning and critical thinking and reflection opportunities before and after each iteration, performed by the consortium of CSO and university partners. Phase 1 involved the elicitation of a preliminary, draft mini-framework of *Values* from the consortium member CSOs. This draft framework provided an initial basis on which the first set of values-based 'indicators' could be built — effectively taking the place of a theoretically-derived framework of values, as outlined in Section 2. It was accepted that this draft framework would be limited, with respect to a 'full' framework of values that might come about from further large and comprehensive studies, but it was hoped that it would provide a workable starting point.

It is important to note that according to some literature (e.g. (Hinkel, 2011)), the term 'indicator' is defined as the *function* linking an observable or measurable variable with a theoretical variable that cannot be directly measured. The values-based 'indicators' created through the ESDinds project (as described in Burford et al., 2013a, 2013b) are thus more correctly understood as *Proto-Indicators*: they are statements or criteria which can guide the identification of measurable indicators for intangible values, through local conceptualization and operationalization.

In Phase 2, the Values and Proto-Indicators were presented to a variety of 'user' CSO project teams for consideration and exploration in the field. The central question was: were these Proto-Indicators useful for developing measurable indicators to assess relevant values-based practices and achievements of the CSOs? And could the results derived be considered valid for that purpose?

Both phases contained multiple iterative consultations and dialogues. In particular it was vital in Phase 1 that the Proto-Indicators were derived from the draft Values framework in a rigorous manner, as they could otherwise endanger the overall validity of any assessments made downstream. A brief overview of that complex piece of work is given below; further detail can be found elsewhere (Podger et al., 2011, 2010). In Phase 2 the emphasis was on the exploration of the *practical utility* of the Values and Proto-Indicators at ground level, with their translation into practice leading back to critical thought about the draft frameworks. This paper is focused on one of those explorations — Echeri — as it revealed significant lessons along the way.

Phase 3 of the work involved taking revised Values and Proto-Indicators to a wider range of CSO projects, to test the validity of results derived from them (e.g. generalizability and further aspects of consequential validity). These results are discussed in Section 5.

3.2. Peer-elicitation of a draft values framework — Phase 1

In the first phase, which we termed *values elicitation*, the consortium aimed to identify (i) words and phrases representing ethical values associated by the CSOs with the 'progress' or 'success' of their projects; and (ii) concrete examples of the enactment of those values in practice, which we viewed as Proto-Indicators — i.e. potential prototypes for measurable indicators of those values (Burford et al., 2013a, 2013b; Podger et al., 2011). This was achieved through a rigorous content analysis of semi-structured interview transcripts, focus group transcripts and key documents (e.g. websites and annual reports). An initial code book was populated with value schemes and classifications from literature, and this was iteratively updated as each researcher identified new themes in the data.

As the analysis yielded 125 values and a very large number of potential indicators, an initial prioritization of values was conducted on the basis of coding frequency — ensuring that the 'top' Value for each CSO was represented. The indicators for the five selected Values (Integrity, Trustworthiness, Unity In Diversity, Empowerment and Justice) were prioritized on the basis of perceived relevance and usefulness by a consensus of CSO representatives within a workshop setting, resulting in an initial list of 177 draft Proto-Indicators taken forward for field testing (sample provided in Table 1).

3.3. Field exploration of values and indicators — Phase 2

The Phase 1 framework of six Values and 177 Proto-Indicators now needed to be explored in Phase 2 to determine if they were relevant, practicable, and able to generate measurable indicators leading to valid assessments. This was to be determined via application by CSOs in their projects. A critical point was whether assessment methods could be found to operationalize the Indicators; they would be useless otherwise.

One of the first CSOs involved in this exploratory work was Echeri, which was chosen because of the willingness of the staff to participate deeply, and the offer of the personal involvement of a co-Director who had experience with non-formal education and creativity. This was thought to be particularly useful for the challenge of developing appropriate assessment methods, and details of the action co-research which took place are presented more fully elsewhere (Podger et al., 2013). Several other pilot tests of the six Values and 177 Proto-Indicators were undertaken with other CSOs who used them in a straightforward manner to develop measurable indicators and assessment tools, and these are reported elsewhere (Burford et al., 2013b). The Echeri study is particularly illuminating and rich with respect to issues of validity and frameworks of values and their relevance to non-formal education, and for that reason is described in some detail in this paper.

3.3.1. Setting

Echeri is a small civil society organization based in Mexico, with a primary focus on reforestation and environmental education through creative arts activities with youth and children. The indicators which their funders and stakeholders commonly use relate to the respective numbers of children involved, educational sessions delivered and trees planted. Projects were typically informally evaluated by collective reflection by team members on the objectives, context, resources, results, foreseen and unforeseen

Table 1
Sample Proto-Indicators for each of the six values identified by participating CSOs.

Value	Sample proto-indicators
Care and respect for the community of life	• The environment/community of life is celebrated • Entity acts to reduce its environmental impact or remedy its contribution to environmental problems • Action is consciously taken to contribute to a greater understanding of the natural world as a source of personal fulfilment • Long term commitments to protect the environment are created and adhered to
Empowerment	• People feel that they are encouraged to reach their potential • People feel that they are provided with opportunities for personal growth • People feel a sense of power that they can effect change • Action is consciously taken to encourage people to express their opinions
Integrity	• People investigate what is right and good by themselves, rather than adopting other people's opinions • People follow through on their commitments • Goals are reviewed between committed parties to determine what has and has not been achieved • Actions of individuals are consistent and in harmony with the core principles promoted by the entity
Justice	• People feel that they are treated equitably and with fairness • Entity acts in a manner that is impartial and non-discriminatory (not discriminating on the basis of nationality, ethnic origin, colour, gender, sexual orientation, creed or religion) • People participate actively in making decisions about issues that affect their lives • People feel that they have an equal opportunity to express their opinions • Decision-making takes into account the social, economic and environmental needs of future generations
Trust/ trustworthiness	• People feel that they are trusted to follow through on their commitments • Entity is transparent about the processes and outcomes of decision-making • Trusted partners are given flexibility to do things differently within prescribed structure • Conflict resolution leads to learning and growth • People are perceived to be truthful
Unity in diversity	• Everyone has their place in the team • Teams include members with different characteristics (e.g. gender, culture, age and other aspects of individual difference such as personality) • Different points of view are heard and incorporated • People learn freely together, regardless of nationality, ethnic origin, skin colour, gender, sexual orientation, creed or religion • People are inclusive (talk to everyone and no one is left out)

obstacles, impact, and proposals for follow-up. Two separate Echeri projects were chosen for the pilot trials of the ESDinds system, based on the interest and availability of the local Director, and the timing of the visit of the university researcher.

The first was a project aimed at reconnecting school children (aged 9–13 years) to their local ecosystem, and equipping them with the attitudes and capabilities to serve as custodians of their local environment. Workshops were carried out in schools, each with 40–60 children, using creative and visual arts, dance, physical movement and theatre. Activities included guided reflection on learning, e.g. about their own local ecosystem and what it contained, followed by technical workshops on tree planting. The children then planted seeds in their own tree nurseries, tended them and took part in tree planting and reforestation on school

grounds and/or in deforested parts of their local community. This programme worked with 15 local schools which had partial indigenous populations, and covered about 15% of that age population within the district.

The second Echeri project was a multi-cultural group of 19 youth aged 12–21, including indigenous and non-indigenous, urban and rural, local and immigrant members from Latin America and Europe. The youth group had been meeting weekly for a year, pursuing a similar non-formal educational programme to the school children mentioned above, but in more depth and scope (e.g. organizing large reforestation campaigns and municipality-wide arts festivals).

3.3.2. Building rapport

Although Echeri was willing to be involved in co-research, the Director was initially sceptical that indicators could be developed for values, and anxious about potential imposition of a numerical scale to quantify the presence of values in her organization. These concerns were addressed through e-mail conversations in which the research team highlighted the action research approach, the focus on local utility, and the specific Values that had been identified in Phase 1 (Trust, Integrity, Justice, Unity in Diversity, Care & Respect for the Community of Life, and Empowerment). The Director felt these resonated with Echeri values, and agreed to pilot the ESDinds system as a 'critical friend' and co-developer. Before the researcher arrived, she worked with the youth group to reflect deeply on the six Values in their local context.

3.3.3. Deciding which values to assess

Organizations usually have several values underpinning their work, but the prototype ESDinds system had only been developed with six. As a first step, Echeri representatives were asked to consider each i) for general relevance to their work and ii) for consideration in the upcoming particular assessment study. No set method was required by the ESDinds system to do this.

The Echeri Director personally interpreted the six Values in the context of her organization, resulting in intriguing differences from the consensus interpretations of the ESDinds consortium. In particular, the Director saw the term Empowerment as closely linked to human rights, invoking a sense of "confronting the *status quo*" which contradicted Echeri's vision of a peaceful transition to sustainability. For this and other reasons, Empowerment was not deemed a value of high priority for assessment in Echeri. The Director also did not see Trust as a high priority for assessment as it was viewed as an outcome of activities, rather than a core value.

The remaining four Values were then taken to a focus group with the members of the Echeri youth group to determine which were 'core'. Each Value was written on one quadrant of a flip chart, and with the Director absent, the youth documented their own definitions and real-life examples of each value in each quadrant. The Director then returned and facilitated the identification of relationships between pairs of values: horizontally, vertically, diagonally and finally spirally. The spiral is a strong indigenous symbol and concept in that locality, and collective decisions were taken to place certain values 'at the centre of the spiral' – on the grounds of a shared understanding that their presence was a prerequisite for the other values to emerge. Through this process, 'Respect and Care for the Community of Life' and 'Collaboration in Diversity' were identified as the most important values for assessment. This decision was the first of several which ensured direct relevance of the assessment, not only to the day-to-day realities of CSO project work, but also to underlying indigenous epistemologies and axiologies.

3.3.4. Deciding on proto-indicators, measurable indicators, and assessment tools

An academic researcher visited Echeri to work with the organisation on the next task: to consider the *Proto-Indicators* associated with each of the two chosen Values, and to determine whether (i) the Proto-Indicators were relevant to them, and (ii) there were any that they would like to be able to assess. Considerations as to which specific *measurable indicators* and *assessment tools* could and would be used were intended to be delayed until after this stage.

The Director and researcher first focused on Proto-Indicators for one Value (Collaboration in Diversity) within the youth group project. The university researcher had been involved in Phase 1 of developing the Proto-Indicators list (in English) and was also a native Spanish speaker, so in principle translation was not problematic from either a conceptual or linguistic point of view. However, what consistently happened was that the Director preferred to not only translate the Proto-Indicators, but to adapt them slightly so that their contextualization within Echeri projects was clear. For example, the indicator #23 was modified from, "The project's activities/events have an emotional effect on participants" to "The project's activities/events produce an emotional connection to the community of life in participants". These 'localizations' did not change the meaning of any Proto-Indicator sufficiently that it could be confused with any other, (i.e. maintaining item validity), but it did mean that the Proto-Indicator was no longer 'standardized' in a reference framework allowing it to be strictly compared to those assessed from the same list for other organizations (endangering generalizability aspects of validity). Because the primary imperative of the research was to develop a tool with practical utility for user CSOs, the ESDinds team decided it was appropriate to 'localize' the Proto-Indicators in this specific pilot project. This would not always be the case, however, as some CSOs might be primarily interested in comparing their values-related achievements with those of other organizations and would therefore require a standardized Indicator set.

During this engaged consideration of the Indicators, the Director occasionally discussed possible *measurable indicators* and methods of assessing them. She finally declared all twelve Proto-Indicators associated with 'Collaboration in Diversity' to be relevant, and set out to assess them all in the youth group project. This choice, again, contributed to the relevance of the assessment; the Director determined which to use.

The researcher then interactively presented 35 variations on several assessment tools, using a handbook previously prepared. The Director produced her own analysis of those she thought were generally suitable for each Indicator. The two then collaborated to develop specific measurable indicators and assessment tools suited to the project. Several types and variations were used, but none in the original form presented – confirming that localization is not merely useful, but key to ensuring relevance and validity. (See striking examples below under Spatial Survey and Word Elicitation.)

The assessment activities were then carried out with the Echeri youth group (results in Table 2a), after which the Director considered which measurable indicators and assessment tools would be appropriate for work in the Schools programme involving 1500 children in 15 schools, often in groups of 60 at a time. She also wanted to develop assessment tools for a second Value – Respect and Care for the Community of Life. Ten of these were chosen for assessment.

Illustrations of the resulting different types of assessment tools and their application to Echeri projects are discussed in turn below. A summary of them, and of the results of each assessment, is given in Table 2. They are emphasized in this paper because they exemplify operationalizing the Proto-Indicators. It was necessary to find

Table 2a
Indicators of "Collaboration in Diversity" assessed through spatial survey and focus group discussions: Echeri youth group.

Echeri localized indicator	Number responding "very much"	Number responding "more or less"	Number responding "not at all"	Comments
Members feel that group norms exist and that they abide by these norms	0	12	0	Immediate response from all members: no hesitation. This indicator triggered animated focus group discussions and transformational learning: youth realised that even though they had set the group norms (e.g. arriving on time, updating the blog) themselves, they were still not fully adhering to them.
Members feel the group is inclusive, with everyone talking to everyone else and no-one being left out	5	7	0	The focus group discussion around these two indicators helped the project director to identify individuals who felt that there were issues with unity and inclusion within the group, and reflect on ways to help them to feel more included.
Members feel the working environment in the group is unified	8	4	0	
Members feel that learning processes in the group accommodate their own learning styles	7	5	0	The focus group helped the project director to identify individuals who felt less certain that their learning styles were being accommodated, and reflect on ways to modify the group's approach in order to serve them better.
Members feel that different points of view are heard and incorporated into actions of the group	6	6	0	Facilitators understood the results as encouraging feedback on their own performance. The focus group discussion around these indicators helped the project director to draw out specific challenges experienced by individual group members in learning together, sharing and being heard.
Members feel they can learn together, share skills, abilities and information freely with one another, regardless of creed, colour, ethnicity, gender	8	4	0	
Members feel that, in participating in the group's vision and activities, they maintain their values and beliefs	11	1	0	In the focus group, the member who answered 'more or less' clarified that he did not feel inhibited or driven into compromise; rather, he felt that his beliefs were in a state of flux so the question of 'maintaining' them was less relevant to him.
Members feel that girls/women are valued in the group	12	0	0	Immediate response from all members: no hesitation. In the focus group, youth realized that the project has successfully generated a space of gender equity, in which (in contrast to national and regional norms) women and men have equal access to information and decision-making. The project director had been working consciously to create this space of equity, but never discussed it explicitly with the group.
Members feel the working environment in the group is harmonious and pleasant	12	0	0	Immediate response from all members: no hesitation. The program director interpreted this as confirmation of the overall validity of the CSO's distinctive methodologies.
Members feel that everyone has their place in the team	12	0	0	
Members feel their individual identity is respected in the group	12	0	0	
Members feel that in the group they are encouraged to reach their potential	12	0	0	

measurable indicators and practicable assessment tools (or combinations) to deliver valid interpretations to the CSOs for their chosen Proto-Indicators. Otherwise the link of validity back to the Value would be broken, and although measurements might be obtained, it would not be clear what they were for.

3.3.5. Applications of assessment tools
3.3.5.1. Spatial and corporal surveys.
Each of the Collaboration in Diversity indicators was converted into a question with a three-point scale. In keeping with the frequent use of spirals as symbols and tools in their regular activities, and the use of long coloured scarves for many of their exercises, a large three-coloured spiral was formed. Each colour represented a point in the scale, e.g. blue being "a lot", green being "more or less", and purple being "a little". After each question, the youth were invited to stand on the colour that most reflected their personal response to the question, e.g. the Proto-Indicator, "Women believe they are valued" became the question: "Do you feel that women are valued in the Echeri youth group?" and in this case all the youth went to the colour showing they felt women were valued. Results are presented in Table 2a.

As the group was small, it was possible to record the responses of individual youth (assisted by video recording), as well as the aggregated results. The results were even richer than expected because it was also possible to note the speed of response; some youth moved without hesitation while others took time to reflect. Similar surveys used corporal methods i.e. body movements such as standing or sitting to denote the answers. Acknowledging the need to explore the underlying reasons for responses, these surveys were followed up with focus group discussions, contributing to validation through triangulation (Gorard and Taylor, 2004).

3.3.5.2. Focus group.
Echeri used the focus group as a follow up to the quantitative spatial survey used to assess Collaboration in Diversity indicators. The focus group concentrated on exploring differences in answers, to gain a deeper sense of both what the indicators meant to the youth, and how they related to them and to one another's answers. For example, for the question, "Do you feel that different points of view are heard and incorporated into the actions of Echeri (A little/more or less/a lot)?" 6 youth felt they were 'a lot', while 6 others felt they were only 'more or less'. This

provided the starting point for a focus group discussion, giving a much richer and subtler picture than the survey alone. It also contributed to the group's own learning and reflection process.

This particular pair of tools — surveys and focus groups — was felt to provide sufficient relevant and valid information to Echeri to allow assessment of current progress relating to one of the Values and its associated Proto-Indicators from the ESDinds system, thus contributing to the validation of the tool.

3.3.5.3. Word elicitation — hand painting circle. This tool was used in the assessment of the indicator, "The project's activities/events produce an emotional connection to the community of life in participants". In this case, the coordinator felt that she knew that the reforestation campaigns had a significant emotional effect on the youth, but wanted to have a more textured understanding of that effect. Accordingly, a word elicitation exercise was developed and adapted to a local activity: the hand-painting circle. The youth were asked to sit in a circle and offer their left hand as a canvas to their neighbour, and paint on their other neighbour's extended left hand. They were asked to paint 'how they felt' at the end of a reforestation campaign in the neighbouring mountains, and when the paintings were completed, to explain what their painting meant to them, while the evaluators recorded the emotional vocabulary used.

The result was a brief list of emotional words for each individual, and an aggregated list providing a group perspective. Words like happiness, joy, and tiredness were common, but there were also individual responses like harmony, wonder, pride, and a sense of contributing to the good of the world.

3.3.5.4. Theatrical comprehension test. This method was used for the Proto-Indicator "Quality of process and results of activities or projects aiming to achieve or promote environmental sustainability". The coordinator wanted to measure the effectiveness of her pedagogical methods in imparting technical information. A comprehension test was devised (detailed in Table 2b) based on local ecosystem awareness and technical knowledge, with emphasis on group rather than individual success.

A single question relating to local knowledge was simply done verbally by asking the group to identify what types of trees grew locally, and noting when mistakes were made and corrected by other members of the group. Most youth knew all the correct answers, and all the youth knew most of the correct answers. A general eco-systemic knowledge was tested by role-playing, through Image Theatre, each of the phases in the natural cycle that creates and maintains a forest, and asking the group to identify them.

Finally, technical knowledge was tested using Forum Theatre, where the researcher began the process of role-playing the first phase of tree-planting by scooping the earth the wrong way, and inviting the youth to identify the mistake, and to take over and correct it. Once a youth had stepped forward and role-played their correction, if it was still mistaken, other youth could interrupt. The group demonstrated that they had a very solid grasp of the eight technical steps, the occasional mistakes being minor and swiftly corrected by a chorus of voices.

The overall result of these assessments was evidence that the Echeri pedagogical processes to achieve and promote sustainability were deemed very effective.

3.3.5.5. Key informant interviews. This tool was used to assess the indicator, "Teams include members with different characteristics (e.g. gender, culture, age and other aspects of individual difference such as personality)", by asking the coordinator to identify aspects of diversity in the group based on her personal knowledge from working with them for over one year.

Table 2b
Indicators of "Care and Respect for the Community of Life" assessed through various methods: Echeri youth group.

Echeri localized indicator	Assessment tool(s)	Results
Pedagogical quality of process and results of activities aiming to achieve or promote environmental sustainability (key messages, facts and skills persist after the workshops)	**Group comprehension test** delivered through multiple choice questions, and through variations on Forum Theatre. The test was designed to measure group, rather than individual knowledge. If the majority of the group knew the answers, the goal was considered achieved. 8 members of the group participated. (a) Knowledge of the local flora was tested via multiple choice questions. (Echeri) (b) The 7-stage natural cycle that creates a forest was tested by the researcher and Juatarhu supporters miming each of the stages, and asking the group to guess what they referred to. (c) The 8 steps and skill-sets involved in artificially creating a forest were represented through forum theatre, where the researcher began by enacting each step with mistakes. Youth had to identify the mistakes and re-enact the step with the correct process. Others were asked to spot any mistakes made by their companions and step in to correct them.	6 of the 8 participating members successfully identified all local trees included in the test. The group successfully identified all 7 stages. In the process, the whole group shouted the correct responses, and each step was precisely and accurately represented, demonstrating a very full grasp (born of experience) from every member of the group.
The CSO's activities generate an emotional connection to the community of life	**Word elicitation through a hand painting circle:** each participant was asked to paint on their neighbour's hand the emotions that they experienced at the end of a tree planting campaign, and then to describe these emotions in words. Researchers listed the words used, and also informally noted the tone of voice and non-verbal communication.	All 7 of the participating members mentioned the words *enjoyment, joy* and *happiness*, the majority using three or more synonyms each; 5 expressed physical sensations (wetness, thirst, muddiness) in a joyful way; 4 mentioned their sense of making a contribution to protecting the environment and a service to the world; 2 mentioned satisfaction at seeing trees planted earlier growing in later months or years; 2 mentioned feelings of harmony; 1 mentioned a sense of wonder and magic.

For the schools programme, the ESDinds system was used to assess which activities supported the conscious aim of contributing to a specific environmental goal adopted by the school, and the level of completion of those goals. It also was used to assess how many workshops had been imparted by Echeri in what schools and with what regularity, towards the indicator, "Education is undertaken to raise awareness and capabilities for the organization to act according to principles of environmental sustainability".

Finally, a key informant interview with a regional expert was used to assess the technical quality of Echeri's reforestation work (see Table 2c).

3.3.5.6. Case study methodology. Given that the Echeri project involved multiple visits to each of 15 separate schools, some of the assessment tools such as key informant interviews were not realistic for all schools, within available timescales and resources. However, the Echeri Director had noticed that the schools fell into three categories of apparent commitment — marginal, moderate and enthusiastic. It was thus decided to carry out 1–2 case studies of schools in each category, to better understand what made one school more committed and receptive than another and the contribution of Echeri activities to this outcome.

3.3.5.7. Indirect measures. Indirect measures were used to assess the indicator, "Long term commitments to protect the environment are created and adhered to." As each participating school had made a long term commitment to create a tree-nursery, an indirect measure of that commitment was its current state. In some schools

the children looked after the nursery well; in others it was littered, or the shelter was torn, etc. Eight schools were surveyed.

Indirect measures were also used with the Proto-Indicator, "Quality of process and results of activities or projects aiming to achieve or promote environmental sustainability". The theatrical comprehension test had measured this indicator in relation to the pedagogical content of the workshops; now, indirect measures were used to measure the technical quality of the actual tree-planting processes facilitated by Echeri in schools. This involved:

- Comparing the number of participating children with the number of seeded bags deposited in the tree nursery. In most visits the number was equal, showing that the workshop was technically effective in producing tree sowing.
- Comparing the number of seeded bags, with the number of germinations obtained.
- Comparing the number of germinations to the number of successful shoots
- Comparing the number of successful shoots transplanted, to the number of trees surviving (by revisiting reforestation sites and counting the surviving trees).
- Comparing all the figures obtained to typical figures expected, as provided by a government tree-planting expert obtained through a key informant interview.

The conclusion from these particular assessments was that, while the pedagogical processes were of high quality, the technical quality of the tree-planting processes was well below the baseline suggested as reasonable by the expert. This was due in part to the

Table 2c
Examples of assessments through various methods in the Echeri Schools Program.

Echeri's localized indicator	Assessment tool(s)	Results
Motivational quality of process and results of activities aiming to achieve or promote environmental sustainability (how well are the tree nurseries looked after and how many children who attended the workshops and/or their families actively participate in looking after them?) Long-term commitments to protect the environment are created and adhered to The CSO's activities generate an emotional connection to the community of life	**Semi-structured observation of tangible project outputs**: the project director, another staff member and an independent observer examined the tree nurseries created by 15 participating schools, noting the overall condition of each nursery and checking for any signs of vandalism or neglect (e.g. litter, damage to infrastructure, growth of weeds). Each nursery was given an overall rating of 'excellent', 'suboptimal' or 'poor' (through consensus of at least 2 observers). **Spatial survey**: The playground of one school was divided into three spaces, representing the answers A LOT, MORE OR LESS, and A LITTLE. Children were asked to move into the space that best represented their answer to the question "How much would you like to look after the tree nursery, and help every day to water the seeds, pick up the rubbish around it, and take out the weeds?"	13 out of 15 tree nurseries were judged to be in excellent condition, 1 suboptimal condition, and 1 poor condition. Among those in excellent condition, one school had completed an entire cycle from tree nursery preparation, through seed sowing, to planting the saplings in a degraded area of land, and had already initiated a second cycle of seed sowing. In the school judged as 'poor condition', the tree nursery had been ripped open in many places, and the soil left unweeded. 57 children answered 'a lot' and 1 'more or less' (but note high probability of social desirability response bias, given that the children were in the presence of Echeri's staff and their own teachers)
Technical quality of process and results of activities aiming to achieve or promote environmental sustainability	**Structured observation of tangible project outputs:** survival rates of trees planted by participating schools **Key informant interview:** the region's foremost tree planting expert was asked to comment on the observed tree survival rates	Survival rates ranged from 18/500 trees planted to 700/1000. In a single case it was zero, as a road was built over the reforested land. The average was 161/500 trees planted. The expert suggested the wide variability was due to differences in suitability of the land for the seeds provided. (With the seeds and resources available, the technique being used was judged to be of satisfactory quality.) The result of this analysis was strategic discussion to explore redefining programme design to give Echeri a larger role in determining the choice of seeds.
Education is undertaken to raise awareness and capabilities for the organisation to act according to principles of environmental sustainability	**Document analysis**: checking reports of work completed to date in order to determine the number of workshops imparted to each school.	Of the 15 schools, all had received at least one workshop (awareness of sustainability and local ecosystems), and the majority had received two or more. It was anticipated to follow this up in future with qualitative case-studies to see how far the themes imparted in the workshops were followed up by teachers and pupils.

unavailability of more appropriate land, and in part to the choice of trees, which were in turn a consequence of funders' constraints as well as a disconnect between forestry experts and community development experts. This new understanding proved very valuable, and is likely to transform the future design of the tree-planting programme.

4. Results

4.1. Interpreting the assessments

Tables 2a and 2b show interpretations of some of the results. The Director in our case study reported that she felt all the Proto-Indicators were capable of being operationalized as measurable indicators, and subsequently of being assessed, even though some methods would need reviewing to eliminate minor problems. For example, there was a high probability of social desirability bias affecting results when children were asked to run to one of three designated spots in the playground representing different answers to the question, "How would you like to look after the tree nursery…" (see Table 2c).

Furthermore, the Director reported that the results were useful, and, in her opinion, allowed valid interpretations as to the status of various values-based outcomes of Echeri work related to the Values of Collaboration in Diversity, and Care & Respect.

For the ESDinds research team, these findings provided evidence that the Indicator system met its requirements of being relevant and validly linked from the measurement level to the values level. The third requirement, of practicality, was also met, as the Echeri Director felt that the assessment activities had been practical and appropriate for them; she had already begun incorporating similar ones into ongoing Echeri activities.

4.2. Issues of validity for the ESDinds indicators

A sub-framework of six Values and 177 Proto-Indicators had been brought to this CSO in the hope that they could be used to provide valid assessments of values-based outcomes – in this case for two grassroots projects.

In order for valid assessments to be made, every link in the chain from the final interpretation, to measurement, to the values framework needed to be considered for validity. First, were the Values considered by the CSO to be relevant to their work, and thus something which was worth measuring to represent successful outcomes? The answer was yes, the evidence being that the Director accepted two of the Values for this purpose, supported by other members of her CSO – the youth. A caveat is that some localized interpretation of the Values meanings took place.

Secondly, were the Proto-Indicators and measurable indicators considered acceptable manifestations of those individual Values – and would measurements of them be appropriate proxies – for the specified use of monitoring outcomes? Again, the answer was yes, because the Echeri Director and other members actively chose (and discarded) Proto-Indicators from the list provided, with those requirements in mind, and when asked, stated that they felt they had sufficient indicators for each Value, with no obvious gaps in coverage. However, they did localize the original Proto-Indicators, making them more directly applicable to the local situations while developing the measurable indicators, which was an important contributor to making their local interpretation very clear and definitely representative of something they wished to measure. The modifications were not sufficient to change one Proto-Indicator nearer to another (i.e. threaten item validity) but very effective in ensuring consequential validity.

Lastly, could assessment tools be found which maintained the chain of validity, i.e. by providing results which the CSOs accepted as valid for the use required, i.e. monitoring and enhancing values-based outcomes? This is more complex to answer because sometimes several measurements with different assessment tools were deemed necessary to fully assess an indicator, e.g. for triangulation purposes. But sometimes the CSO was only interested in an approximate or partial assessment of a specific Proto-Indicator, in light of their overall purpose. Ultimately, the CSO had to play the main role of determining the overall consequential validity of the results obtained, by considering their overall appropriateness for the intended use. And in this work, Echeri representatives deemed their results to have consequential validity, reporting satisfaction that the results allowed valid monitoring interpretations as required, and without gaps, thus confirming that the ESDinds Values and Proto-Indicators were successful for this purpose. For the ESDinds team, this provided evidence that the evaluation system was successful in its intended job of facilitating good consequential validity.

5. Discussion

Having established that the ESDinds values-focused evaluation system was successful for Echeri and other pilot applications, it is appropriate to carefully consider wider issues. A key one is comparability. If each organization uniquely develops measurable indicators, and uses individualized assessment tools to assess them, then one ends up comparing spatial survey results, observation notes and written questionnaire results as different assessments of, e.g. the practice of gender equality. We would argue that this is not an impossible situation, and conversions could be used to normalize them, e.g. a 'traffic light' colour coding in which red, yellow and green respectively represent poor, satisfactory and good practice in relation to predetermined standards for each indicator (c.f. Milman and Short, 2008). We also suggest that the overall validity of such comparisons would likely be better, as the alternative would be rigid, highly defined indicators requiring specific prescribed assessments – which would then have far less local relevance and thus consequential validity. Even then, after translation into other languages and application in different cultures, it is unlikely that the final interpretation would be consistent. Evidence for this is provided by a recent study of human values across 25 countries through the European Social Survey, in which it was demonstrated that the invariance of theoretical value constructs across countries and over time cannot be assumed (Davidov, 2010). We thus argue that the localization allowed in the ESDinds system should be seen as a strength, rather than a weakness: it is a very important departure from traditional assessment methods, and one which may be the key to success in non-formal EE/ESD and in other arenas of environmental management.

Phase 3 of the ESDinds project included applications with a wider range of CSOs, including the use of the ESDinds system for evaluations of activities of a youth theatre group in Germany, a project run by the Red Cross in Sierra Leone, and an Earth Charter workshop for university students (Burford et al., 2013b). The provision of ESDinds as a user-friendly framework, which was not so rigid as to deny localization, proved useful in each case. The ESDinds system has also been put onto a web platform, www.wevalue.org, where it has been used by scores of diverse CSOs from around the world. Although it has wide applicability, there is clearly room for improvement, as expected from any first-generation tool.

It may be remembered that in the absence of a useful pre-existent values framework, we devised our own in Phase 1,

reducing it to the 'top six' Values to keep it manageable in the field. It was surprising that all of the CSOs piloting the ESDinds system, and scores who later worked with it online, found at least one Value of this small subset to be of key interest. This suggests that any 'full' framework of Values for this kind of assessment purpose might be small – possibly only requiring 10–12 values. Thus, a small amount of further work might yield a significant values framework with a very wide range of applicability e.g. across protected area management, conventional biodiversity conservation programs, adaptive co-management initiatives, public-private partnerships and community-based organizations.

In addition, the Echeri Director's interpretation of the ESDinds Values as non-equivalent (e.g. with Respect as fundamental to all of them, at the 'centre of the spiral', and Trust seen as a result of the others) contradicts existing Western academic values frameworks which set them out as semi-equivalent items which cluster together. Indeed, the Echeri Director's interpretation of the value-labels diverged greatly from those intended by the researchers. Both of these differences raise serious questions about the validity of traditional values assessment systems based only on ranking or rating of single value-labels (e.g. Rokeach, 1973, 1979; Schwartz, 1992, 1994), although some concepts from social psychology literature – such as Rokeach's distinction between instrumental and terminal values- may still be useful to CSOs.

Furthermore, when reflecting on their assessment results, the Echeri Director commented that she felt the Value of 'Empowerment' had been covered to some extent, so that, in her judgement, using the Proto-Indicators from the other two Values could also provide valid partial assessment for this third Value. In fact, after considerable further discussions with the Director and other members Echeri it was understood that Echeri would have preferred having the pool of all 177 Proto-Indicators available for them to make their own value-indicator links. This has since been done in later pilots.

If the Proto-Indicators are 'uncoupled' from their original value-labels, it becomes necessary for local stakeholders to make their own intersubjective judgements about which theoretical variable(s) are 'indicated' by the chosen measurable variables (cf. Hinkel, 2011). In the pilots, this question has been addressed in two different ways. Some organizations have linked the ESDinds Proto-Indicators to their own established values vocabulary – using them, for example, to measure 'Participation' or 'Innovative Thinking' as locally understood (Burford et al., 2013a). Others have used the ESDinds system to obtain useful assessments of values-related outcomes *without* linking them locally to specific value-labels, e.g. by measuring 'pro-success values' for non-formal EE/ESD projects in a more general sense (WeValue, 2013).

Another point of note from this study is that CSO members using the ESDinds system usually undergo a 'Eureka moment' – i.e. a significant moment of 'transformational learning' or group realization, when they suddenly find they can collectively understand how to conceptualize shared approaches that previously had not been articulated. This happens as they read through the list of Indicators for those relevant to them. It is possible that discussions of each item help in building a new shared vocabulary to describe concepts already present but not articulated, and suddenly the group seems to see a 'bigger picture' or landscape of their work, in new terms. At this point, they have sometimes decided their mission statements need urgent revision. This effect suggests new ideas for theories on the relationships between personal inner values and those operationalized into action, which we expect to document more fully elsewhere.

In summary, the Echeri study suggests significant variations and contradictions to existing values frameworks and assessment systems, directly challenging some concepts in values theories. We

expect to devise new work in these areas and invite other interested parties to collaborate with us.

Our findings suggest that valid values based assessments can be made, thus bringing values into the realm of evaluation. Such evaluations could take different forms, according to the requirements of CSOs and their funders. They might, for example, involve comparing post-intervention data with pre-intervention baselines; monitoring project implementation processes; comparing outcomes obtained by different organizations, or different interventions within the same project; or monitoring against other variables, e.g. monetary cost or time. Funders might specify their own choice of Proto-Indicators, or ask CSOs to nominate some for negotiation. Although the Proto-Indicators have been designed to be localisable by the CSOs, measurable indicators could be firmly set for evaluation once a consensus has been reached with funders.

While this paper has focused on evaluating projects rather than improving them, it is worth noting that the ESDinds system also has the potential to catalyse positive change. We have elsewhere demonstrated important 'process use benefits', i.e. beneficial outcomes deriving from the actual process of participating in the evaluation, irrespective of its findings (Burford et al., 2013b; see also Forss et al., 2002) The explicitly values-based nature of ESDinds can assist stakeholders to conceptualize and communicate their espoused values more clearly, and as our project results have indicated, the ESDinds indicators may help to promote environment- and ethics-centred discussions in contexts where these concerns are normally overlooked (ESDinds 2011). This became evident when the ESDinds system was used within a private-sector financial services company with no prior orientation towards EE or ESD. A senior executive of the company reported that through discussing ESDinds indicators with the CEO, he had realised for the first time that the environmental issues which mattered to him as an individual could also constitute valid topics for professional strategy formulation (WeValue, 2013).

5.1. Implications for environmental management

Beyond the specific use of the ESDinds framework described above, there is also significant potential for broader application. While this paper has focused specifically on educational components, we can envisage that the existing indicator set could be usefully applied to the holistic evaluation of large environmental management programs. The need for values-based approaches to both project design and project evaluation becomes evident when we reflect on findings such as those of Fielding et al. (2013), who describe empirical tests of interventions aimed at reducing water consumption among Australian households. While all the interventions were successful in the short term, water usage returned to pre-intervention levels just twelve months after the termination of the project. Although a complex situation, we speculate that if indicators relating to "sense of power to effect change" and "emotional connection to the community of life" had been incorporated from the start, both the project design and its long-term impact might have looked different.

In seeking new applications for our framework we acknowledge, however, that the existing set of values-based Proto-Indicators is not comprehensive. It is likely that new Proto-Indicators would need to be added in order to provide adequate evaluations of ambitious environmental management initiatives such as National Parks or large-scale community-based conservation programmes. In designing new proto-indicator sets or measurable indicators from scratch, it would be important to identify all relevant groups of stakeholders, making particular efforts to involve those whose voices often go unheard. (In the Echeri evaluation, for example, the

Indigenous youth participated actively in co-defining and choosing the values to be assessed). This may require some creativity in research design, and careful separate consideration of the depth, breadth and scope of participation in each case (Harder et al., 2013).

The principles underlying ESDinds also have the potential to contribute to wider environmental management applications beyond the arena of project evaluation. These include environmental valuation, environmental assessment, and in particular the design and use of sustainability indicators/sustainable development indicators (SDIs) at national and global levels. There are three specific elements of ESDinds which significantly extend the current discourse on participatory development of SDIs (Bell and Morse, 2008; Bell et al., 2012; Gudmundsson, 2003; Rametsteiner et al., 2011; Rosenström and Kyllönen, 2007). One of these elements is the ESDinds project's explicit focus on ethical values, which are arguably of such central importance to sustainable development that they should be regarded as a key element of a fourth 'pillar' of sustainability alongside environmental, social and economic aspects (Burford et al., 2013a). Another element is the peer elicitation/ peer validation method used to derive the Proto-Indicator list, which, as we discuss elsewhere, simultaneously reassures users with familiar vocabulary and challenges them with new ideas (Burford et al., 2013b). Finally, the third principle is the localizability and context-sensitivity of the ESDinds system, which may be critically important. Both peer-elicitation and localizability resonate strongly with the emerging focus on participation, recently acknowledged as fundamental to the very nature of sustainable development (Bell et al., 2012).

While the 'global quest for indicators' is often misinterpreted as a 'quest for global indicators', the ESDinds project provides a model for alternative polycentric approaches (c.f. Ostrom, 2010). By building in an ethical values focus, peer-elicitation and localizability, indicator developers can honour and accommodate very diverse world-views and understandings of sustainability. These three principles — which have the potential to contribute to radical transformation of the ways in which indicators are conceptualized, developed and used — are particularly timely within the context of ongoing debates about the Sustainable Development Goals (SDGs) intended to succeed the Millennium Development Goals in 2015. If, as asserted by Forss et al. (2002), the benefits of evaluation derive not only from the findings but also from participation in the process of defining indicators and tools, there is an urgent need to reflect on values and participation within the SDG creation process. We have developed this concept further elsewhere (Burford et al., 2013a).

6. Conclusion

In this paper, we have described the development and application of ESDinds, a values-based system for evaluating non-formal environmental education and ESD initiatives. The more significant problems of assessment in non-formal EE/ESD are overcome, including non-traditional, multiple and multi-level goals, alienation by external vocabulary and concepts, and lack of local relevance. The particular problem of undeveloped goal conceptualization is overcome by the use of the CSO peer-elicited lists, which act not only as a selection list 'menu' but also as a prompt — stimulating participants to articulate and reconsider their own true goals in a collective way, catalysed by those of the contributing peer organizations as reflected in the Proto-Indicator list.

The ESDinds system appears to provide a useful first step towards an 'effectiveness revolution' in environmental management (c.f. Keene and Pullin, 2011) in that it (a) successfully integrates top-down and bottom-up perspectives, and (b) permits the identification and operationalization of values-related processes and soft

outcomes, previously regarded as intangible and immeasurable. The underlying principles of localizability, peer elicitation/peer validation and the use of an ethical values lens all contribute to making ESDinds an important model for sustainability assessment in a more general sense, especially in the context of post-MDG global target setting.

Acknowledgements

References

Allport, G.W., Vernon, P.E., Lindzey, G., 1951. A Study of Values. Houghton Mifflin, Boston, MA.

American Educational Research Association, 1999. The Standards for Educational and Psychological Testing. Developed jointly by American Educational Research Association (AERA), American Psychological Association (APA) and National Council for Measurement in Education (NCME). AERA Publications, Washington, D.C.

Anello, E., November 2006. Ethical Infrastructure for Good Governance in the Public Pharmaceutical Sector: Working Draft for Field Testing and Revision. World Health Organization, Department of Medicines Policy and Standards (PSM), Geneva, Switzerland.

Bell, S., Morse, S., 2008. Sustainability Indicators: Measuring the Immeasurable? Earthscan, London, UK.

Bell, S., Morse, S., Shah, R.A., 2012. Understanding stakeholder participation in research as part of sustainable development. J. Environ. Manag. 101, 13–22.

Bryman, A., 2007. Barriers to Integrating Quantitative and Qualitative Research. J. Mixed Methods Res. 1, 8–22.

Burford, G., Hoover, E., Velasco, I., Janouskova, S., Jimenez, A., Piggot, G., Podger, D., Harder, M.K., 2013a. Bringing the 'missing pillar' into sustainable development goals: towards intersubjective values-based indicators. Sustainability 5, 3035–3059.

Burford, G., Kissmann, S., Rosado-May, F.J., Alvarado Dzul, S.H., Harder, M.K., 2012. Indigenous participation in intercultural education: learning from Mexico and Tanzania. Ecol. Soc. 17, 33.

Burford, G., Velasco, I., Janouskova, S., Zahradnik, M., Hak, T., Podger, D., Piggot, G., Harder, M.K., 2013b. Field trials of a novel toolkit for evaluating 'intangible' values-related dimensions of projects. Eval. Program Plann. 36, 1–14.

Cedefop, 2001. European Commission Communication: Making a European area of lifelong learning a reality. COM (2001), 678.

Cameron, K.S., Quinn, R.E., 1999. Diagnosing and Changing Organizational Culture. Based on the Competing Values Framework. Addison — Wesley Longman, Massachusetts.

Carleton-Hug, A., Hug, J.W., 2010. Challenges and opportunities for evaluating environmental education programs. Eval. Program Plann. 33, 159–164.

Christensen, L., Nielsen, J.E., Rogers, C.M., Volkov, B., 2005. Creative data collection in nonformal settings. New. Dir. Eval. 2005, 73–79.

Clavijo, K., Fleming, M.L., Hoermann, E.F., Toal, S.A., Johnson, K., 2005. Evaluation use in nonformal education settings. New. Dir. Eval. 2005, 47–55.

Cohen, L., Manion, L., Morrison, K., 2011. Research Methods in Education, seventh ed. Routledge, Abingdon, UK.

Crishna, B., 2007. Participatory evaluation (I) — sharing lessons from fieldwork in Asia. Child. Care Health Dev. 33, 217–223.

Crohn, K., Birnbaum, M., 2010. Environmental education evaluation: time to reflect, time for change. Eval. Program Plann. 33, 155–158.

Daigneault, P.M., Jacob, S., 2009. Toward accurate measurement of participation rethinking the conceptualization and operationalization of participatory evaluation. Am. J. Eval. 30, 330–348.

Davidov, E., 2010. Testing for comparability of human values across countries and time with the third round of the European Social Survey. Int. J. Comp. Sociol. 51, 171–191.

Donaldson, S.I., Patton, M.Q., Fetterman, D.M., Scriven, M., 2010. The 2009 Claremont debates: the promise and pitfalls of utilization-focused and empowerment evaluation. J. Multidiscip. Eval. 6, 15–57.

Ebrahim, A., 2002. Information struggles: the role of information in the reproduction of NGO-funder relationships. Nonprofit Volun. Sect. Q. 31, 84–114.

Ebrahim, A., 2003. Accountability in practice: Mechanisms for NGOs. World Dev. 31, 813–829.

Edwards, M., Hulme, D., 1996. Too close for comfort? The impact of official aid on nongovernmental organizations. World Dev. 24, 961–973.

Ellis, R., Hogard, E., 2006. The trident: a three-pronged method for evaluating clinical, social and educational innovations. Eval. Int. J. Theory, Res. Pract. 12, 372–383.

Ferraro, P.J., Pattanayak, S.K., 2006. Money for nothing? A call for empirical evaluation of biodiversity conservation investments. PLoS Biol. 4, 0482–0488.

Fielding, K.S., Spinks, A., Russell, S., McCrea, R., Stewart, R., Gardner, J., 2013. An experimental test of voluntary strategies to promote urban water demand management. J. Environ. Manag. 114, 343–351.

Fitzpatrick, J.L., 2012. Commentary—Collaborative evaluation within the larger evaluation context. Eval. Program Plann. 35, 558–563.

Fleming, M.L., Easton, J., 2010. Building environmental educators' evaluation capacity through distance education. Eval. Program Plann. 33, 172–177.

Flowers, A.B., 2010. Blazing an evaluation pathway: lessons learned from applying utilization-focused evaluation to a conservation education program. Eval. Program Plann. 33, 165–171.

Ford Foundation, 2011. Our Approach to Impact Assessment and Learning. Ford Foundation, New York, NY.

Forss, K., Rebien, C., Carlsson, J., 2002. Process use of evaluations: types of use that precede lessons learned and feedback. Evaluation 8, 29–45.

Fraser, E.D.G., Dougill, A.J., Mabee, W.E., Reed, M., McAlpine, P., 2006. Bottom up and top down: analysis of participatory processes for sustainability indicator identification as a pathway to community empowerment and sustainable environmental management. J. Environ. Manag. 78, 114–127.

Gorard, S., Taylor, C., 2004. Combining Methods in Educational and Social Research. Open University Press, Buckingham, UK.

Gramsci, A., 1971. Selections from the Prison Notebooks. Lawrence & Wishart, London.

Gudmundsson, H., 2003. The policy use of environmental indicators − learning from evaluation research. J. Transdiscipl. Environ. Stud. 2, 1–12.

Habermas, J., 1971. Knowledge and Human Interests. Beacon Press, Boston, MA.

Handy, R., 1970. The Measurement of Values. W. H. Green, St Louis.

Harder, M.K., Burford, G., Hoover, E., 2013. What is participation? Design leads the way to a cross-disciplinary framework. Des. Iss. 29 (4), 41–57.

Heimlich, J.E., 2010. Environmental education evaluation: reinterpreting education as a strategy for meeting mission. Eval. Program Plann. 33, 180–185.

Hinkel, J., 2011. "Indicators of vulnerability and adaptive capacity": towards a clarification of the science−policy interface. Glob. Environ. Change 21, 198–208.

Hogard, E., 2008. Purpose and method for the evaluation of interpersonal process in health and social care. Eval. Program Plan. 31, 34–40.

Holte-McKenzie, M., Forde, S., Theobald, S., 2006. Development of a participatory monitoring and evaluation strategy. Eval. Program Plan. 29, 365–376.

Horáková, N., 2005. Co je pro nás v životě důležité? Naše společnost 2.

Keene, M., Pullin, A.S., 2011. Realizing an effectiveness revolution in environmental management. J. Environ. Manag. 92, 2130–2135.

Kluckhohn, C., 1951. Values and value-orientations in the theory of action. In: Parsons, T., Shils, E. (Eds.), Toward a General Theory of Action. Harvard University Press, Cambridge, MA, pp. 388–433.

Kopelman, R.E., Rovenpor, J.L., Guan, M., 2003. The study of values: construction of the fourth edition. J. Vocat. Behav. 62, 203–220.

Kushner, S., Norris, N., 1980. Interpretation, negotiation and validity in naturalistic research. Interchange 11, 26–36.

Lather, P., 1986. Research as praxis. Harv. Educ. Rev. 56, 257–277.

Lincoln, Y.S., Guba, E., 1985. Naturalistic Inquiry. Sage, Beverly Hills, CA.

Lincoln, Y.S., Guba, E., 1986. But is it rigorous? Trustworthiness and authenticity in naturalistic inquiry. In: Williams, D.D. (Ed.), Naturalistic Evaluation. Jossey-Bass, San Francisco, CA, pp. 73–84.

Messick, S., 1989. Validity. In: Linn, R.L. (Ed.), Educational Measurement, third ed. Macmillan, New York, pp. 13–103.

Messick, S., 1995. Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. Am. Psychol. 50, 741–749.

Milman, A., Short, A., 2008. Incorporating resilience into sustainability indicators: an example for the urban water sector. Glob. Environ. Change 18.

Onwuegbuzie, A., Leech, N.L., 2005. On becoming a pragmatic researcher: the importance of combining qualitative and quantitative research methodologies. Int. J Soc. Res. Methodology 8, 375–387.

Onwuegbuzie, A., Witcher, A., Collins, K., Filer, J., Wiedmaier, C.l., Moore, C., 2007. Students' perceptions of characteristics of effective college teachers: a validity study of a teaching evaluation form using a mixed-methods analysis. Am. Educ. Res. J. 44, 113–160.

Ostrom, E., 2010. Polycentric systems for coping with collective action and global environmental change. Global Envir. Change 20, 550–557.

Parker, R., Bradley, L., 2000. Organisational culture in the public sector: evidence from six organisations. Int. J. Public Sect. Manag. 13, 125–141.

Podger, D., Piggot, G., Velasco, I., Hak, T., Janouskova, S., Zahradnik, M., Dahl, A., Burford, G., Hoover, E., Harder, M.K., 2011. ESDinds Working Paper: Eliciting Values and Values-based Indicators. www.esdinds.eu.

Podger, D., Piggot, G., Zahradnik, M., Janouskova, S., Velasco, I., Hak, T., Dahl, A., Jimenez, A., Harder, M.K., 2010. The earth charter and the ESDinds initiative: developing indicators and assessment tools for civil society organisations to examine the values dimensions of sustainability projects. J. Educ. Sustain. Dev. 4, 297–305.

Podger, D., Velasco, I., Amezcua Luna, C., Burford, G., Harder, M.K., 2013. Can values be measured? Significant contributions from a small civil society organisation through action research evaluation. Action Res. 11 (1), 8–30.

Rametsteiner, E., Puelzl, H., Alkan-Olsson, J., Frederiksen, P., 2011. Sustainability indicator development: science or political negotiation? Ecol. Indic. 11, 61–70.

Reed, M., Fraser, E.D.G., Morse, S., Dougill, A.J., 2005. Integrating methods for developing sustainability indicators to facilitate learning and action. Ecol. Soc. 10, 1–6.

Reed, M.S., Fraser, E.D.G., Dougill, A.J., 2006. An adaptive learning process for developing and applying sustainability indicators with local communities. Ecol. Econ. 59, 406–418.

Reino, A., Vadi, M., 2010. What Factors Predict the Values of an Organization and How? University of Tartu, Estonia.

Riddel, R.C., 1999. Evaluating NGO development interventions. In: Lewis, D. (Ed.), International Perspectives on Voluntary Action: Reshaping the Third Sector Earthscan, London, pp. 222–241.

Rode, H., Michelsen, G., 2008. Levels of indicator development for education for sustainable development. Environ. Educ. Res. 14, 19–33.

Rokeach, M., 1973. The Nature of Human Values. The Free Press, New York.

Rokeach, M., 1979. Understanding Human Values. Free Press, New York.

Rosenström, U., Kyllönen, S., 2007. Impacts of a participatory approach to developing national level sustainable development indicators in Finland. J. Environ. Manag. 84, 282–298.

Sagiv, L., Schwartz, S.H., 1995. Value priorities and readiness for out-group social contact. J. Pers. Soc. Psychol. 69, 437–448.

Schlater, J.D., 1969. Investigating values underlying family decisions. Res. Bull. 23.

Schlater, J.D., Sontag, M., 1994. Toward the measurement of human values. Fam. Consum. Sci. Res. J. 23, 4–25.

Schwartz, S.H., 1992. Universals in the content and structure of values: theoretical advances and empirical tests in 20 countries. Adv. Exp. Soc. Psychol. 25, 1–65.

Schwartz, S.H., 1994. Are there universal aspects in the structure and content of human values? J. Soc. Issues. 50, 19–45.

Schwartz, S.H., Melech, G., Lehmann, A., Burgess, S., Harris, M., Owens, V., 2001. Extending the cross-cultural validity of the theory of basic human values with a different method of measurement. J. Cross-Cultural Psychol. 32.

Silver, H., 2003. Does a university have a culture? Stud. High. Educ. 28, 157–169.

Singh, P., 2011. Developing a community of thinking: assessment of environmental education. Environ. Educ. Res. 17, 113–123.

Sleeswijk Visser, F., Stappers, P.J., Van der Lugt, R., Sanders, E.B.-N., 2005. Context mapping: experiences from practice. CoDesign 1, 119–149.

Spranger, E., 1928. Types of Men [Originally Published in German by Niemeyer, Halle (Saale), 1914, as *Lebensformen;* Translated by P. J. W. Pigors]. G. E. Stechert Company, New York.

Springett, J., 2001. Participatory approaches to evaluation in health promotion. In: Rootman, I., G, M., Hyndman, B., McQueen, D.V., Potvin, L., Springett, J., Ziglio, E. (Eds.), Evaluation in Health Promotion: Principles and Perspectives. WHO Regional Publications, Copenhagen, pp. 83–105.

Springett, J., 2003. Issues in participatory evaluation. In: Wallerstein, M.M.N. (Ed.), Community-based Participatory Research for Health. Jossey-Bass, San Francisco, CA.

Steen, M., Manschot, M., De Koning, N., 2011. Benefits of co-design in service design projects. Int. J. Des. 5, 53–60.

Stringer, L.C., Dougill, A.J., 2013. Channelling science into policy: enabling best practices from research on land degradation and sustainable land management in dryland Africa. J. Environ. Manag. 114, 328–335.

Sutherland, W.J., Pullin, A.S., Dolman, P.M., Knight, T.M., 2004. The need for evidence-based conservation. Trends Ecol. Evol. 19, 305–308.

Tashakkori, A., Creswell, J.W., 2007. Exploring the nature of research questions in mixed methods research. J. Mixed Methods Res. 1, 207–211.

UNECE, 2005. Background Paper on Development of Indicators to Measure Implementation of the UNECE Strategy. UNECE.

UNECE, 2006. UNECE Strategy for Education for Sustainable Development, Guidance for Reporting.

Wainwright, D., 1997. Can sociological research Be qualitative, critical and valid? Qual. Rep. 3.

WeValue, 2013. Case Study: Testing Values-based Indicators in a Financial Services Company.

Wightman, F., 2010. Measures of success: not everything that counts can be counted. Vanc. Found. Mag. (Fall 2010), 8–9.

Williams, R.M., 1979. Changes and stability in values and value systems, a sociological perspective. In: Rokeach, M. (Ed.), Understanding Human Values. The Free Press, New York, pp. 15–46.

Wiltz, L.K., 2005. I need a bigger suitcase: the evaluator role in nonformal education. New. Dir. Eval. 2005, 13–28.