



MANUAL ON ESD EVALUATION

MODULE 3 *FORMAL METHODS FOR ESD EVALUATION*

EURETA ROSENBERG
LUDWIG CHANYAU

Module 3

Formal Methods for ESD Evaluation

Eureta Rosenberg
Ludwig Chanyau



Publisher

Eberswalde University for Sustainable Development (HNEE) / Leuphana University Lüneburg

Year

2021

Design

Canva

This Manual is licenced under [Creative Commons Attribution-Non Commercial 3.0 Germany](https://creativecommons.org/licenses/by-nc/3.0/de/) (CC BY-NC 3.0 DE)

The ideas and opinions expressed in the manual are those of the authors; they are not necessarily those of the institutional partners, Engagement Global or of the Ministry for Economic Cooperation and Development (BMZ).

THE FOUR MODULES OF ESD EVALUATION COURSE ARE DESIGNED TO BE USED IN CONJUNCTION WITH EACH OTHER AND HENCE SHOULD NOT BE USED INDEPENDENTLY. MANY OF THE TERMS AND CONCEPTS USED IN THE MODULE ASSUME KNOWLEDGE THAT COME FROM ALL THE SECTIONS OF THE ESD EVALUATION MANUAL.

For queries, please write to esd_evaluation@leuphana.de

SUSTAINABLE DEVELOPMENT GOAL 4

4 QUALITY EDUCATION



Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all

Target 4.7 by 2030 ensure all learners acquire knowledge and skills needed to promote sustainable development, including among others through education for sustainable development and sustainable lifestyles, human rights, gender equality, promotion of a culture of peace and non-violence, global citizenship, and appreciation of cultural diversity and of culture's contribution to sustainable development

ACKNOWLEDGEMENTS

The development of this module would have not been possible without the generous and varied contributions of many individuals and university partners.

Editors

Deepika Joon, Leuphana University of Lüneburg, Germany
Jennifer Krah, Eberswalde University for Sustainable Development, Germany
Ludwig Chanyau, Rhodes University, South Africa

Project Leader

Heike Molitor, Eberswalde University for Sustainable Development, Germany
Jennifer Krah, Eberswalde University for Sustainable Development, Germany

Authors

[ESD and Evaluation](#) by Matthias Barth, Professor at Leuphana University of Lüneburg
[Emergence of Evaluation in ESD](#) by Rob O'Donoghue, Emeritus Professor at Rhodes University South Africa. Contribution on [Regenerative ESD and Evaluation](#) by Heila Lotz-Sisitka, Distinguished Professor at Rhodes University
[Formal Methods for ESD Evaluation](#) by Eureka Rosenberg, Professor at Rhodes University and Ludwig Chanyau, Rhodes University. Contributions on [Quality and Validity for ESD Evaluation](#) by Juan Carlos A. Sandoval Rivera, Researcher at Universidad Veracruzana, Mexico
[Communicating ESD Evaluation](#) by Nadine Shovakar, Member of ESD Expert Network and Lena Hoffmann, Eberswalde University for Sustainable Development

Lead Facilitators of Summer academy

Deepika Joon, Leuphana University of Lüneburg
Ludwig Chanyau, Leuphana University of Lüneburg (DAAD Fellow) and Rhodes University

Team Members

Anwesa Goswami
Annika Bucher
Carolin Ellerkamp
Charline Rieffel
Edali Beltran
Johanna Philipps
Maxi J Baumert
Selena Dami

Moodle Adaptation

Giovanni Fonseca

Merle Dreessen

Milena Weber

Video

Avishek Shovakar

Graphic Recording

Louisa Syzmorek

Partners

ESD Expert Network

Eberswalde University for Sustainable Development, Germany

Leuphana University of Lüneburg, Germany

Rhodes University, South Africa

Universidad Veracruzana, Mexico

The ESD Evaluation course has been brought together after efforts of more than two years. The Working Group on ESD Evaluation commenced its work after the presentation by Rob O'Donoghue and the team at Bonn Conference on "From Knowledge to Action: ESD versus Climate Change" in November 2017. Deepika Joon and Eureka Rosenberg presented the concept of ESD Evaluation at the CIES 2019 conference in San Francisco which was followed by the first meeting of the Working Group in Berlin in August 2019.

We would like to acknowledge with much appreciation the participants of the summer academy on "Regenerative ESD and Evaluation for the Future We Want" held from 21 July to 20 August 2021 for their feedback on the curriculum. Special thanks to Preven Chetty for his work on module 2. The speakers of the summer academy Aymara Llanque Zonta, Deepak Ramola, Heila Lotz-Sisitka, Injairu Kulundu-Bolus, Jennifer Krah, Jiska Troppenz, Louisa Syzmorek, Saransh Sugandh and Thomas Salmon provided critical insights into the topic of ESD and Evaluation. In addition, we would like to thank colleagues from Engagement Global for their support

a VALUE CREATION FRAMEWORK

with
Ludwig Chanayau
& Thomas Salmon

by Etienne Wenger
Beverly Trayner
& Maarten de Laat

ESD EVALUATION
SUMMERSCHOOL
18.08.21



impact Evaluation
during and after
programmes

→ realist Evaluation
looks at underlying
programme mechanisms



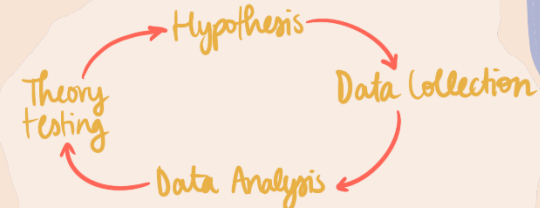
Evaluation is
'normative'!

people are
evaluative!

what works
for whom in
what
circumstances?

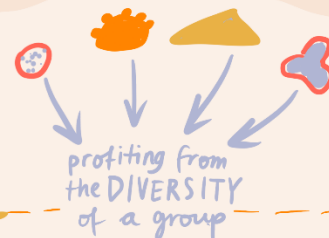


what
you ASSUME
determines
what you FIND!



THEORY of CHANGE ANALYSIS
understanding the HOW and WHY

a Social
Learning
space:



mutual
ENGAGEMENT

common
ENDEAVOUR

shared
REPertoire

commitment
mutual
trust

shared
reframing



individual + collective
value creation

finding
EVIDENCE



you might end up
having a lot of data



how to collect it?
e.g. Atlas
T.I.

in the end it's about
Validating your data

VCF

strategic value

measures Value
in different cycles:

Immediate value
potential value
applied VALUE
Realised value
transformative value
enabling value

orienting value



Module 3

Formal Methods for ESD Evaluation

3.1 BROAD LEARNING OBJECTIVES	1
3.2 INTRODUCTION TO THE CONTENT AND STRUCTURE OF THE CONTENT IN MODULE 3	2
3A INTRODUCTION TO METHODS & APPROACHES.....	4
3A.1 FORMAL METHODS AND ESD EVALUATION	4
3A.2 TERMINOLOGY.....	6
3A.3 BROAD APPROACHES TO EVALUATION.....	15
3B FORMAL METHODS FOR EVALUATING ESD.....	25
3B.1 EMPIRICIST APPROACHES	25
3B.2 CONSTRUCTIVIST APPROACHES TO EVALUATION	29
3B.3 PROCESS-BASED EVALUATIONS	39
3B.4 THEORY-DRIVEN APPROACHES	44
3B.5 PARTICIPATORY APPROACHES	57
3B.6 COMPLEXITY-SENSITIVE APPROACHES	67
3C CONCLUSION AND CONNECTION WITH OTHER MODULES OF ESD EVALUATION CURRICULUM	72
3D ASSIGNMENTS.....	74
3E QUALITY AND VALIDITY IN ESD EVALUATION (SUB SECTION OF MODULE 3)	77
SLIDES USED IN THE PRESENTATION ON INTRODUCTION TO FORMAL METHODS FOR ESD EVALUATION.....	85
TRANSCRIPTION OF DIALOGUE ON QUALITY AND VALIDITY FOR ESD EVALUATION.....	97

Module 3

Formal Methods for ESD Evaluation

Eureta Rosenberg & Ludwig Chanyau

3.1 BROAD LEARNING OBJECTIVES

After the completion of the module, it is expected that the participants are able to:



... Develop an understanding of a range of relevant formal methods for evaluating ESD activities and programmes, and the broader approaches associated with these methods.



... Learn the main features, underlying assumptions, value and limitations of the evaluation methods (methodology) in relation to ESD evaluation.



... Implement well-reasoned, informed decisions about suitable evaluation methods for a particular ESD evaluation purpose and context, and source further information required to use these methods for planning a formal ESD evaluation process.



... Design a formal ESD evaluation process by applying their learning in the module.

3.2 INTRODUCTION TO THE CONTENT AND STRUCTURE OF THE CONTENT IN MODULE 3


This is the third module in a short course in Education for Sustainable Development (ESD) Evaluation for ESD practitioners in diverse contexts including ESD centers, schools, universities, organizations, and multi-institutional programmes. Module 3 is a substantial module in two parts, with 3A being more broadly orientating and 3B being more specific.

Evaluation courses and handbooks abound, and yet there remains huge interest among managers and practitioners in how to undertake evaluative processes that are specifically suitable for transformative approaches to education and learning. A clear case can be made for field specific evaluation competencies, in this case ESD specific evaluation competencies, which both embrace evaluation theory, and critique evaluation theory, on those aspects where it fails to align with or support ESD evaluation as introduced in Modules 1 and 2.

The learning objectives will be achieved by engaging with the content of the two-part module, which will first briefly introduce some terminology, followed by a spectrum of evaluation approaches and their underlying assumptions (Module 3A). Module 3B zooms in on a selection of methods exemplifying different approaches, and unpacks methods in terms of their key features, underlying assumptions, value and limitations, and the kinds of contexts and purposes for which they have been used. Additional readings illuminating both the theory and the practical application of the methods should also be perused for a thorough understanding of the methods and associated approaches and how to use them practically.

There are several in-course activities and two course assignments associated with Module 3 (both A and B together). In assignment 3.1 students will work in pairs or groups, to analyze and then demonstrate (through roleplay or other means), two different ESD evaluation methods, comparing and contrasting them. Course participants will be asked to work with others to encourage the formation of a learning network; many evaluation processes require teams working together. Feedback on assignment 3.1 should inform assignment 3.2.

In Assignment 3.2 students individually consider a particular ESD context of their choice and a particular evaluation purpose and decide which method(s) could be used in such a context and for such a purpose.



They then design an ESD related evaluation for that context and purpose, to more deeply explore the use of a particular evaluation approach and method (or combination of approaches and methods). (See below for explanation of terminology.) In certain formats of the course, following Module 4, students may be required and supported to implement their design; reflect on the outcomes; and use the experience to re-design the evaluation process to apply, deepen and consolidate their learning. This will prepare them for conducting a formal evaluation in the ESD field.

3A INTRODUCTION TO METHODS & APPROACHES

3A.1 Formal Methods and ESD Evaluation

In Module 2, evaluation was introduced as an approach and a process that is integral to ESD: Education for Sustainable Development is an evaluative undertaking, as learners and educators together interrogate the status quo, and ask if and how it needs to change. Global visions like the SDGs and local experiences and inquiries can form the basis for such evaluative processes as an integral part of ESD (ESD as evaluation of current practices and the status quo). Educators and the learners themselves can also engage in evaluating their learning support processes and their learning towards sustainability (evaluation in ESD). Is it therefore necessary to dedicate time to think separately about evaluation methods, rather than ESD methods?

It is! From time to time, there is a need to more formally gather information about ESD processes and engage in interpreting that information with others, so as to gain more distanced perspectives (zooming out). That is, a more formal evaluation is needed. ESD funders and institutions may want a formal account of what has been achieved through the ESD activities, and ESD practitioners themselves may want to use a formal evaluation report for communications purposes. These formal evaluations can be undertaken by the ESD practitioners and learners themselves, or external evaluators can assist. In these cases, there is an expectation and need to use an evaluation design that is professionally recognized as legitimate. But even this formal method should be appropriate for the transformative learning approach that is 'education for sustainable development'.

This module takes that intention into account and helps evaluators to choose or adapt a formally recognized method that can also be fit for purpose.

Wenger-Trayner and Wenger Trayner (2020) argue that learning to make a difference (for example ESD) is a form of social learning, in which people engage with each other in collectives, develop a sense of what change they want to bring about, develop the understanding needed to make that difference, and pay attention to new information arising from their learning and change actions. The four connected modes of social learning are: creating value, translating that value into something useful, evaluating the value, and

framing the value. This theory of social learning is one of many theories about learning, education and development, that includes an explicit focus on evaluation as integral to learning (of individuals and collectives, such as organisations). This makes the evaluation of learning a special case of evaluation: it has to evaluate an inherently evaluative activity, and in itself, the evaluation itself also has to be a source and a process of learning!

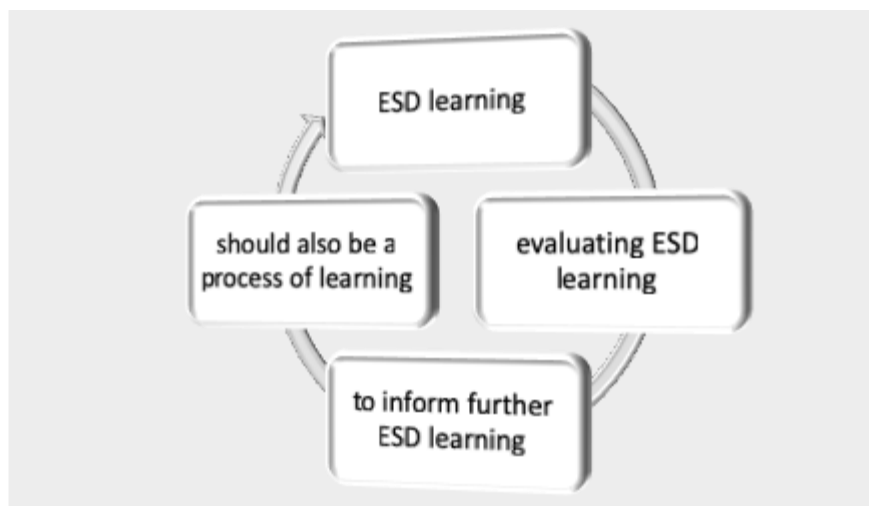


Figure 1: The special case of evaluating learning programmes in ways that support more learning

As one examines a formally recognized evaluation method (such as a Most Significant Change (Story) method) for its suitability for ESD processes, one therefore needs to consider how it will serve a multi-purpose goal: evaluation that is accountable to funders and institutions, that can produce credible evaluation outcomes to communicate to external parties, and at the same time, is a source of learning for all involved. The same consideration was applied when the next section, on Terminology, was compiled. How are standard M&E terms applicable in this special context of evaluation in transformative learning processes, like ESD?

3A.2 Terminology

When engaging with the literature on evaluation, with other evaluators and evaluation participants, it is helpful to know basic terms so as to plan and implement well-informed evaluation processes with others and demonstrate their credibility when necessary. It is also important, at the start of this module, to agree on how some key terms will be used. In the field and in the literature the same term (e.g., outcome) can be used in different ways, and different terms (e.g., method, approach, process) can be used to refer to more or less the same thing. Some method-specific terms are elaborated in Module 3B. Here in Module 3A, the focus is on some generic or commonly used terms. The chosen description is adapted to the ESD context.

Methods, approaches, instruments, tools

These terms are very generic and can be used in a variety of ways; here is an explanation of how they are to be used in this module:

An **evaluation method** is a set of guidelines and decisions to make on how to design and do a particular type of evaluation, which will unfold differently from another type of evaluation. The method (also referred to as the evaluation design) involves a coherent set of considerations, that will guide a formal, coherent and defensible evaluation process to address a certain purpose while meeting certain expectations from the stakeholders. An evaluation method is similar to a research method, where an experimental method for example, is different to a narrative enquiry method. Guidelines for how to design and do a narrative enquiry evaluation and how to design and do an experimental evaluation will differ, each set of guidelines will be coherent and characteristic (albeit with some overlap), and so these are two different methods.

The term **approach** is used to mean a broader set of considerations than methods. For example, one may take a broad participatory approach, within which one would then choose one of several possible participatory methods, e.g., a participatory action research, or a participatory appreciative enquiry. A method can also be described as an approach, depending on the level of specificity one has in mind, e.g., one could consider a case study approach (broadly referring to assumptions about using cases rather than other units such as whole populations) or a case study method, which would refer to more specific guidelines on how to go about evaluating a case.

To conduct an evaluation process, with its characteristic approach and method, one may require the use of some tools, activities or instruments. E.g., the method of 'most significant change (stories)' uses a narrative approach, as well as a participatory approach, in which activities like story solicitation and tools like story selection templates could be used. Tools, activities and instruments are terms often used interchangeably, with the difference being the extent to which they are structured. For a large-scale survey, for example, a tightly structured instrument (the questionnaire or list of



questions, exactly as they are to be asked) is used, so that all survey participants respond to the same questions, and answers can be aggregated for statistical comparisons. The evaluator is not likely to change the instrument (questionnaire) mid-way through the survey. In a workshop, on the other hand, activities would be planned beforehand, but also flexible enough (i.e. less structured) to be adjusted if necessary as the workshop unfolds, to better engage the participants towards the desired purpose. The term tools can be used quite generically to refer to a wide range of online and printed reporting formats and templates, question sets, question designs (such as the well-known Likert scale), statistical techniques and formulae for calculating differences.

Many of these will be encountered through Module 3, and there is merit to consider tools, activities and instruments that have been tried and tested, but evaluators can also design their own.

Monitoring and Evaluation

Development agencies and government funded initiatives often use the terms monitoring and evaluation (M&E) together, as a system for measuring changes in programmes (or policies) and assessing their impact (OECD, 2011, p.7). Monitoring in these contexts involves the recording of activities and results, that will then be collated to account for how resources have been used, and to what end. Monitoring provides the information needed for accountability purposes, but also data that can be used for evaluation purposes (see below).

Monitoring can also be done by programme participants (or citizen scientists and community activists) to monitor what is happening to Commons resources like rivers, air quality and forests, to pick up the impacts of pollution for example, and track improvements brought about by sustainability actions. Citizen monitoring is a powerful approach to ESD, another example of the evaluative nature of ESD.

In Module 3, we focus on evaluation methods, but we can assume that in longer-term evaluation systems, monitoring will be part of a coherent monitoring-and-evaluation system. Generically, monitoring is simply data collection, that requires a protocol, agreement on what is to be monitored, why, how, how often, by whom, and how the information should be stored and shared. In specific evaluation methods, monitoring practices will be influenced by the method, e.g., one will collect stories or quantitative data; it will be collected by external parties or programme participants. When evaluations are undertaken, the quality and relevance of the monitoring data should also be evaluated and adjustments made where necessary.

Evaluation is needed to complement monitoring, for while monitoring can alert us to problems that may be arising, or improvements, evaluation is needed to understand the nature and causes of the problem, and the reasons behind the improvement. Molund and Schill (2007, p.12) noted that, while monitoring provides records of activities and results, and signals problems to be remedied along the way, it may not be able to explain why a particular problem has arisen, or why a particular outcome has occurred or failed to occur. To deal with such questions of cause and effect, an evaluation may be required. Only with deeper understanding can we decide how to address a problem, or under what conditions a successful intervention can be scaled out, or scaled up.

Molund and Schill (2007, p.11-12) describe an evaluation as a reality test, a learning mechanism that provides feedback on the results of action in relation to programme objectives, plans, expectations or standards of performance. The requirement that evaluations should be systematic and credible derives directly from the function of evaluation as a reality test. To serve as a probe on the realism and accuracy of plans and expectations, an evaluation must use sound and transparent methods of observation and analysis (ibid).

There are many kinds of evaluations, that serve different purposes in the life of a programme: from diagnostic evaluations and feasibility assessments, to design evaluation, implementation evaluation,

economic evaluation (like cost-benefit analyses), outcomes evaluation and impact evaluation, as well as synthesis or meta-evaluations that collate and analyze the findings of other evaluations.

Assessment and Review

An assessment is a critical evaluation of information, for purposes of guiding decisions on a complex, public issue. The topic is defined by the stakeholders, who are typically decision makers. Assessments are conducted by a credible group of experts with a broad range of disciplinary experience, in a balanced and transparent way. Assessments reduce complexity but add value by summarization, synthesis and sorting what is known and widely accepted, from what is not known (or not agreed). Assessments relate to the situation at a particular time and in a given geographical domain. They are often repeated after some period. Table 1 is based on the explication of scientific assessments in the report on the Millennium Ecosystem Assessment (2005). (www.millenniumassessment.org) (Source: PAGE, 2016, Green Economy Learning Assessment South Africa, DEA, DHET, UNITAR and Rhodes University, https://www.un-page.org/files/public/green_economy_learning_assessment_south_africa.pdf.)

Table 3.1 Comparing Reviews with Assessments

	Review	Assessment
Audience	Scientists	Decision-makers
Done by	One or a few	Large and varied group
Topic	Simple and narrow	Broad and complex
Identifies gaps in	Research: curiosity driven	Knowledge for implementation: problem-driven
(Un)certainty statements	Hidden	Required, but clearly flagged
Coverage	Exhaustive, historical	Sufficient to deal with main range of uncertainty
Synthesis	Not required	Essential to reduce complexity

Effectiveness and efficiency

Effectiveness is about whether a programme achieves what people want it to achieve. In the context of interventions that have clearly defined, predetermined targets or goals, effectiveness refers to whether the

targets or goals have been met as a result of the implementation of planned activities. In the first case we are concerned with the achievement of targets for the production of goods and services (such as ESD materials or workshops), in the second with the achievement of the further effects that we intend to bring about through these materials and workshops. Effectiveness at the output level is no guarantee for effectiveness in terms of outcomes and impacts. An intervention may achieve all its targets with regard to goods and services, and still not be effective at the outcome and impact levels (Molund & Schill, 2004, p. 28). Effectiveness should ideally be determined at the level of outcome and impact.

Efficiency is about what lies between the process of intervention and the results. More exactly, it is the ratio of the value of the results of an intervention to the value of the resources used to produce them. An intervention is optimally efficient if its value is greater than the value of any alternative use of these resources. If the same resources could have produced better results in some other way, or if the same results could have been produced with fewer resources, it is less than fully efficient (Molund & Schill, 2004, p.37).

Inputs, outputs, activities vs outcomes and impact

Inputs in ESD initiatives are all the resources that are required for the ESD programme or process being evaluated, that which is being used to do the work; often considered to be what donors, governments and other parties invest to make the programme happen. This usually includes staff, physical resources, teaching materials, and funds. Very often, inputs can also be outputs, that is, if practitioners produce what they need themselves, as part of the ESD process. Examples of inputs that can also be outputs (and then become inputs again in further processes) include policies, partnership contracts or formal agreements, and teaching materials.

Activities are the work being done in ESD programmes, using the inputs and producing outputs and hopefully also outcomes (results) in the process. In ESD programmes the activities can be workshops, lectures, field trips, networking, practical experiments, modeling, forecasting, tree planting, gardening, building, producing new policy, producing teaching materials, doing radio and TV broadcasts, and more.

Outputs are materials and other products produced and services delivered, e.g., a funded programme may deliver a series of ESD books or workshops which would then be called its outputs (workshops could also be called activities); while outputs and activities are easy to count up, they are seldom the final intended outcomes of a programme.

Inputs and outputs language is associated with simple engineered systems (think of a pipeline) and so-called results-based and public management. That is why it is often difficult to apply these concepts to more open-ended, complex and emergent systems and processes.

Outcomes are results or intended benefits, in the form of specific changes, “what we wish to achieve”. They can be divided into initial (short term), intermediate and long term and they are the results of the interactions between the programme processes, inputs, activities and outputs, as well as contextual factors, foreseen and unforeseen. There can be much disagreement about what constitutes an outcome vs an output, and also about the difference between outcomes and impacts. In some classification systems, the impact is the more immediate result, with the outcome being the longer term, and ultimately desired result. In other classifications, it is the opposite.

Impacts are, in the system used here (based e.g., on SIDA’s Evaluation Manual, Molund and Schill, 2007) usually regarded as the longer term or ‘final’ results of achieving all the desired outcomes. The impact is the operationalized version of the vision that has been set for the programme at its start, although there is also a recognition that impacts can be both intended and unintended, both positive and negative. The difference between outcomes and impacts is often one of scale (e.g., individual outcomes, system level impacts) and/or time (outcomes in the shorter term and eventual impacts in the long term).

ESD example:

The impact we would like to see from a decade of ESD, for example, would be citizens, industries and governments all using planetary resources in a socially just and on an ecologically sustainable basis. Working back from this ultimate impact, the long-term outcomes of a decade of ESD would be sustainability and justice being considered by citizens, industries and governments in all their decisions and actions. Working back from there, more short-term outcomes would be that these role-players have a good understanding of the nature of and need for sustainable development. When we think about how that is to

be achieved, we think about activities and outputs (like knowledge resources and processes for building agency) and finally, about the tangible resources we will need for these ESD activities, i.e. budgets, staff, skills, materials, and so on.

Indicators

Indicators are the observable occurrences which provide evidence that something significant has happened – whether an output was delivered, an intermediate outcome occurred or long-term changes are manifest (Bakewell, et al. 2003, p.21).

The simplest definition is a “piece of information that provides evidence of a change”. Indicators can be quantitative (e.g., the number of new ESD courses introduced at universities in 2020) or qualitative, requiring a more subjective evaluation. It is imperative to mention that indicators provide an idea of the changes that took place, but do not explain the change process. It is also important to ensure that the means of collecting indicators will remain constant over time. This is crucial in order to ensure comparability, especially in the context of adaptation where end of programme evaluations may not take place until twenty or fifty years after programme completion (Lamhauge, et al. 2012, p.24).

Programme, Project, Intervention

Programme is a set of interrelated interventions, projects or projects that are typically implemented over an extended period of time, often by several parties, and may cut across sectors, themes and/or geographic areas. A programme may have a range of strategies working towards defined outcomes, for example it may be a mixture of development, relief, advocacy, networking and capacity building (Office of the Director of U.S. Foreign Assistance, 2009, p.9). In an ESD context, a programme is a coherently planned set of ESD activities or interventions over a longer period of time.

Project is a discrete activity (or ‘intervention’) implemented by a defined set of implementers and designed to achieve specific objectives within specified resources and implementation schedules. A set of projects make up the portfolio of a programme (Office of the Director of U.S. Foreign Assistance, 2009). At the heart of project design is deciding what to improve or change, and how a project can bring about that change.

However not all projects bring about dramatic change. Stabilizing a situation or slowing the rate of decline could be just as important (IUCN, 2004, p.6). In an ESD context, a project might be a discrete activity, such as an experimental planting or a short-term workshop series on a discrete topic.

Intervention is an action or entity that is introduced into a system to achieve some predetermined outcome or result. In the programme evaluation context, an intervention refers to an activity, project or programme that is introduced or changed, amended, expanded, etc. (Office of the Director of U.S. Foreign Assistance, 2009, p.9). ESD as an educational process is not usually conceptualized as an intervention. However, a donor or government might propose an intervention to strengthen a country or organisation's capacity to undertake ESD.

Accountability

Sida (Molund and Schill, 2004, p.14) defines accountability as the relationship which exists where one party – the principal – has delegated tasks to a second party – the agent – and the latter is required to report back to the former about the implementation and results of those tasks. If an ESD agency has received funding from someone such as SIDA to undertake ESD activities, then that form of accountability will apply, based on the agreement between the parties (such as a contract) and often evident in reporting.

However, accountability also involves the ESD agent's answerability vis-à-vis other partners, such as the ESD programme participants and intended beneficiaries, like teachers and learners. In this case reporting is still important, but can take a variety of different forms, as discussed in Module 5 of this course. ESD participants and stakeholders can be accountable to each other on a range of issues, not just the use of funding. Some distinguish between financial accountability, which is answerability for the allocation, disbursement and utilization of funds, and performance accountability, which concerns results. Evaluation, in so far as it serves as a tool for accountability, provides information for reporting about performance and results. It is less concerned with financial accountability, which is mainly the province of auditors and accountants.

Note: The explanations have been kept short here; for a deeper understanding, including a sense of where there are different uses of the same terms, and how different terms are shaped by the assumptions

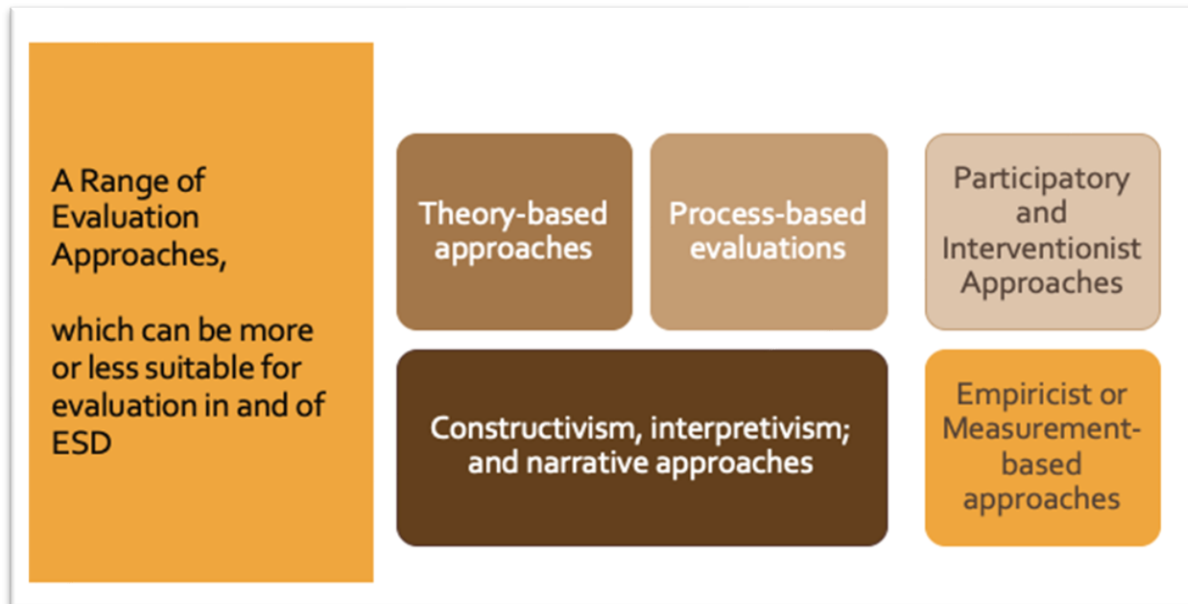
underpinning the particular evaluation approaches in which they are most frequently used, please refer to the references below.

References

- Bakewell, O., Adams, J. and Pratt, B. (2003). Sharpening The Development Process: A Practical Guide To Monitoring And Evaluation (Praxis Guides).
- Funnel, S.C. and Rogers, P.J. (2011). Chapter 2: Variations of program theory over time, pp. 27-30. In: Purposeful Program Theory. Effective use of theories of change and logic models. Jossey-Bass, Wiley, New York.
- Giffen, J. (2009). The Challenges of Monitoring and Evaluating Programmes.
- IUCN (The World Conservation Union) (2004). Core Concepts in Planning, Monitoring and Evaluation of Projects in IUCN (Global M&E Initiative).
- <https://www.cbd.int/doc/pa/tools/Core%20Concepts%20in%20Planning,%20Monitoring%20and%20Evaluation%20of%20Project%20in%20IUCN.pdf>
- Lamhauge, N., E. Lanzi and S. Agrawala (2012). Monitoring and Evaluation for Adaptation: Lessons from Development Co-operation Agencies. OECD Environment Working Papers, No. 38, OECD, Paris. <http://dx.doi.org/10.1787/5kg20mj6c2bw-en>
- Molund, S., and G. Schill (2007.) Swedish International Development Cooperation Agency (SIDA), Department for Evaluation and Internal Audit. Looking Back, Moving Forward: SIDA Evaluation Manual (2nd revised edition).
- OECD (2011). Section 10: Monitoring and Evaluation, in The OECD DAC Handbook on Security System Reform: Supporting Security and Justice. OECD. https://www.oecd.org/dac/conflict-fragility-resilience/publications/ssr_section%2010.pdf

3A.3 Broad Approaches to Evaluation

Table 1: Map of different approaches to evaluation methodology



There are many ways in which to classify different approaches to evaluation. Evaluation typologies can vary according to (for example) ...

- **contexts** (e.g., in national policy contexts, or international development intervention contexts, or health programmes, educational programmes, social work programmes, etc.)
- **purposes** (e.g., in a government policy evaluation context, different evaluation types can be described as diagnostic evaluations, feasibility studies, design evaluations (e.g., strategy evaluations), implementation evaluations, economic, impact and synthesis evaluations...
- or **philosophical and methodological framings** (approaches to knowledge and how best to produce it).

In this course we use the third framing. We focus largely on programme-based contexts, in education and development. A note on scope: We will not be extending discussions about economic evaluations (such as cost-benefit or efficiency evaluations).

Empiricist Approach to Evaluation

Evaluation methods typically used with an empiricist approach include experiments (test group(s) compared to a control group), quasi-experimental designs (e.g., pre-test post-test comparisons), surveys and some systematic reviews. These methods can also be used with non-empiricist approaches. So, what exactly is meant by empiricism?

An empiricist approach to evaluation privileges measurement, in the form of quantitative data that can be statistically manipulated, to the exclusion of all other forms of information. It is based on the premise that one can only trust that which can be mathematically measured, because everything else is too subjective to present the truth about the matter. Empiricism is closely aligned with a logical-positivist approach to science, and with what philosophers like Santos (2018) call 'modern Western science' in its narrowest sense.

This approach has strong support among governments and donors, and has a long history in evaluation, and so it is easy to find information on how to design such an evaluation, and it can be readily communicated. By cutting out most of the contextual variables in a situation, and just focusing on a small set of variables of interest, which are turned into measurables that can be statistically analysed, empiricist approaches create the possibility of a straight-forward design and easy-to-communicate findings.

The challenge with an empiricist approach is that it has to try and create laboratory-like conditions, which on its own, does not accurately represent reality and as a stand-alone, is not that helpful for evaluation of and learning in ESD. One is then also required to leave out a lot of contextual and other relevant factors which may not be amenable to measurement or manipulation. This approach represents what Bhaskar (2010) called a shallow ontology, that is, important layers of social reality are not taken into account in the knowledge production (in this case evaluation) process.

This does not mean that methods like surveys and experiments are not useful if one wants to avoid the shallow or narrow nature of empiricism. They can be used combined with other methods, with other underlying approaches. The empiricist underpinnings - that only that which can be measured, actually exists - would first have to be discarded. For example, one could do a survey of ESD practices at universities

across the country in a participatory manner, with academics, students and operational staff first coming together to decide what might be important to include in the survey, and also following up the completion of the survey with focus group discussions at the various universities to discuss the findings in more depth, in relation to findings from elsewhere. This offers evaluation the 'best of both methods' (case studies and surveys) and is often described as a 'mixed methods' approach. For a non-empiricist approach to systematic reviews, Pawson and other realist evaluators are helpful resources (see www.betterevaluation.org).

Constructivist Approaches to Evaluation

In contrast to the experimental sciences which seek to discover immutable causal laws and relationships, interpretivism approaches the analysis of human actions as enquiry in search of meaning. Constructivism is related to interpretivism and takes the view that knowledge is created or constructed, not discovered, as the product of complex, discursive practices.

Researchers do not discover truths about social life or education; they build models and concepts, often by working with others, to make sense of it, and test and modify these in the light of further experience. Thus, people's interpretations, opinions, perceptions, cultural values and institutional norms, are all regarded as valuable information in the quest for understanding what is going on, and whether a social undertaking like ESD is worthwhile or not. Constructivism is a general philosophy (an epistemology and, in the case of radical constructivism, also an ontology) that has informed education, ESD and evaluation. In the field of evaluation, the so-called Fourth Generation Evaluation Approach (Guba and Lincoln, 1989) has been influential as an interpretivist and constructivist approach.

It is also worth noting, evaluation approaches based on the premise of radical constructivism, which has been associated with post-structuralism and the assumption that all of human reality is socially and personally constructed. This ties in complex ways to postmodern approaches to research and evaluation, where the search for a universal or even a particular truth, is abandoned. Stronach and McLure (1997), writing from a postmodern perspective, argued that despite the radically open nature of reality, useful evaluations can nonetheless be undertaken. They explain how in an article "Can the Mothers of Invention Make Virtue out of Necessity?" which reviewed the common practice of "quick and dirty" evaluations.

Not all constructivist approaches to evaluation assume a radical constructivist ontology; most assume a realist ontology (Bhaskar, 2010), in which there are layers of reality beyond individuals' interpretation of them, and mechanisms that are social in nature but nonetheless quite intransient, so that some statements and conclusions will be 'truer' and more reality-congruent than others. Subjectivity is addressed through 'inter-subjective' objectivity and multiple data sources and methods are used to give a fuller, more dimensional picture of reality. This is sometimes referred to as triangulation, with either the one truth, or the one best version of the truth to arrive at through the different data sources.

Sometimes constructivist approaches are used by an external evaluation team to arrive at a conclusion about a programme towards the end of a programme's life, which is then reported to the programme stakeholders. At other times, there is an emphasis on evaluation throughout the programme's life, with a view to also continuously improve the programme as it unfolds.

Process-based evaluations are not epistemologically different from constructivist evaluations, but to highlight the process dimension, they are introduced here as an additional category; they could also be seen as a sub-category of constructivist evaluation, and they are often used to respond to the complex nature of social settings, as described later in the Module in complexity-sensitive approaches.

Process-based Approaches

These are evaluations that recognize the value of ongoing evaluative thinking and evaluative processes, extending from evaluations that are only done at the end of a programme or mid-way through a programme, to evaluations that are done as part of programme activity, by the practitioners or programme implementers themselves, with or without other stakeholders, or by external evaluators, or a combination of these role-players. Patton (2011) has promoted and illuminated the notion of developmental evaluation as a prime example of process-based approaches to evaluation, in his book that linked it to complexity and the need for ongoing learning (informed by evaluation activities) when rolling out complexity-sensitive programmes.

Process-based evaluations recognize that evaluation is not just about accounting for how resources have been used, or proving whether a programme has been a success or not, but a source of ongoing learning

that can be used to improve the programme. Process-based evaluations also recognize that many educational and social change initiatives involve processes with significant outcomes that might only manifest or be visible far in the future (an example from the natural resource management field is strategic adaptive management), and that in the interim, it is useful to get evaluative feedback so as to steer or redirect the programme, and to recognize whether some worthwhile outcomes can already be celebrated and used to further direct the programme. The value-creation evaluation method (Wenger-Trayner and Wenger-Trayner, 2020) is an example of the latter process-based evaluation.

Participatory & Interventionist Approaches

Participatory evaluations are examples of process-based and developmental evaluations and also a sub-category of constructivist approaches, broadly. They are discussed as an additional category, along with interventionist approaches, because the participatory and interventionist dimensions highlight additional guiding principles and considerations that are worth discussing under a separate heading. Module 2 describes the participatory turn that has been prominent in both ESD and broader evaluation fields.

Participatory approaches are based on the assumption that those who implement programmes and those who are meant to benefit from programmes are not just good sources of information for external evaluators to consult; they are in fact well placed to conduct the evaluations themselves. They usually have a very high stake in whether the programme is successful or not; they are on site most of the time; they have special insider insights but these insights may only surface if they approach the programme with an evaluative lens; and their emerging insights can steer the focus of evaluative processes towards the most pertinent questions that need to be asked at the time. In fact, in some ways anyone contributing to an evaluation is already a participant, and this approach recognizes that and optimizes it.

This relates to the epistemological and political reason for knowledge co-construction; as argued by Santos (2018) and others, the current world presents as a 'cognitive empire' in which only certain (modern Western scientific) knowledge forms are regarded as credible, leaving everyone impoverished, because (according to Santos) all knowledge is partial, and a much wider range of ways of knowing (including modern Western science) is needed to understand (for example) sustainability problems and how to respond to them. Participatory research and evaluation methodologies can help to expand the range of knowledge forms

ESD practitioners and stakeholders have access to (establishing rich 'knowledge ecologies' that compensate for their partial natures), provided the evaluation methodology allows for the recognition of these 'other ways of knowing' (See section on the validation of evaluation findings). These may include the voices of marginalized people(s), knowledge of the body or senses, knowledge lost in undocumented archives, or knowledge born from struggle. As potentially a form of decolonial practice, participatory research must address the insider/outsider dilemma, i.e., the extent to which insiders (or outsiders) have the right to claim evaluative knowledge of a particular situation; and to craft data collection (or data co-generation) methods that are not extractive in nature, and are suitable for bringing to the fore knowledge that had in one way or another been oppressed before.

An interventionist approach is specifically designed to take some action to address a particular issue, and at the same time evaluate the outcomes of the action, through reflective processes, before taking further action. It can be undertaken by one person only but is often participatory in nature as well. Interventionist approaches to evaluation can be an integral part of daily organisational life, e.g., in the case of strategic adaptive management and evaluation, where it is often associated with a desire for ongoing learning and capacity development on the part of the participants.

In which case it would require learning-focused evaluation (Patton, 2011). Alternatively, it could be a once-off or occasional event to address a discrete problem. An example of the latter may be a reflective practitioner evaluation, where an educator, for example, or a group of educators and university researchers, may decide to introduce a new approach to ESD in the classroom, and study the outcomes. When the students themselves become involved in the design and evaluation process, it can be described as a participatory and interventionist process, that is integrally educational, and developmental, and evaluative in nature.

Theory-driven Approaches to Evaluation

Theory-driven approaches are aimed at evaluating societal and institutional intervention programmes, such as the introduction of a new ESD policy for schools, or an intervention to better resource ESD at all universities, or the implementation of a programme to strengthen farmers ability to use new sustainable agriculture practices.

Theory-driven approaches are quite specific to programme evaluation, and here it is necessary to distinguish between evaluating ESD programmes aimed at social or institutional impacts, e.g., introducing or improving ESD processes or outcomes (evaluation of ESD programmes) vs evaluating ESD processes and learning outcomes (evaluation in ESD). The programme and the educational impacts are of course linked, as the nested diagram of EE in, of and as evaluation in Module 2 illustrates. A distinction is however necessary here because we can develop a theory of change for a programme but not always for an educational process or curriculum. The educational logic in a learning programme or curriculum is philosophically different to the change logic of a societal or institutional programme intervention logic. Theory-driven approaches to evaluation apply to programmatic interventions designed to bring about a predefined change: The policy is meant to guide schools on a new curriculum to teach, but it is not the curriculum itself. There can be a drive (programme or other intervention) to strengthen ESD at universities, but the drive/ programme is often not the same as the ESD processes themselves.

So how do we go about evaluating programmes aimed at supporting or enabling ESD, and what are theory-driven approaches about? Surely all programmes are based on theories of how the change they want to see in the world, will come about through this programme? That is indeed so, but in practice, very few programmes are explicit about what exactly this theory is, which makes it difficult to evaluate the programme, unless of course, the evaluation is actually aimed at identifying the theory. In theory-driven approaches to evaluation one may therefore start by surfacing and articulating the programme theory, then designing the evaluation from there. Or, one can intentionally undertake an evaluation (often a meta-evaluation) to develop theory.

A programme theory approach to evaluation aims to explain what happens between the intervention and the eventual outcome, and why this is occurring.

The interest in the theoretical underpinnings of interventions is enjoying a particularly popular moment at this time. In *Purposeful Programme Theory: Effective use of theories of change and logic models*, Funnel and Rogers (2011) traced the history of theory-driven approaches to evaluation all the way back to Kirkpatrick who, in the 1960s, used four domains of change for evaluating training. In more recent times theory-driven evaluations have become popular where programme developers recognized complexity in

the situations they were trying to change, to the extent that there is now talk of it representing a fifth generation of evaluation theory (Brouselle & Buregeya, 2018).

Complexity-sensitive Approaches to Evaluation

The notion that social situations are complex systems is not new, but it has only recently started to receive wider attention in the evaluation community. Based on the disappointing value of evaluation findings compared to the complex questions that need answering in social situations, there have been calls for evaluators to recognize these situations as complex systems. If ESD in higher education, for example, involves complex systems, then evaluation methods that treat them as simple systems are seen to be inadequate for evaluating them. Hence the call for what has come to be termed complexity-sensitive evaluations (Britt, 2016).

Some of the features of complex systems is that they have many dynamically and often unpredictably interacting parts. Whereas complicated systems (like a space rocket) also have many interacting parts, those parts are finite in number (the system is closed) and they interact in predictable ways. Therefore, when complicated systems are evaluated, it is easy to set expected outcomes and associated indicators for them and monitor and evaluate them accordingly.

What makes systems complex as opposed to complicated, is that (1) they are open, hence it is harder or impossible to pin down the number and nature of interacting parts, (2) the interactions between parts are unpredictable, and (3) they are characterized by emergence, that is, surprising outcomes arise from these often-unforeseen interactions, even when there is some intentionality in the system (that is, it is not random).

A complex system is not chaotic and not completely unknowable; much can be learnt (e.g., through evaluation) about what is going on and why, if one tracks the patterns of interactions and looks for more intransient mechanisms that could give rise to the observable events, interactions and outcomes in the system. For example, capability, agency and/or sense of community could be underlying mechanisms giving rise to citizens starting a recycling or reforestation programme. Identifying the presence of these underlying mechanisms could lead to some predictions (Bhaskar, 2010; Pawson and Tilley, 1997).

Complexity-sensitive evaluations respond to the presence of complexity not by trying to reducing the complex situation to something simpler, because then their findings will not apply to the actual situation, but by taking cognizance of the features of complex systems. This means being open to emergence, and to acknowledge that there will always remain a fair amount of uncertainty as to what is actually going on. This approach may not be well received by some evaluation stakeholders, but as Wenger-Trayner and Wenger-Trayner (2020, p.26) put it, “Uncertainty, in this case, is a safeguard against oversimplification, biases, and assumptions. Uncertainty is not the opposite of certainty, but its critical friend. ... Uncertainty is a precaution. It is a way to encompass complexity, to weave it into knowing”. Two recognized methods for evaluation under conditions of complexity are the value-creation framework for evaluating social learning spaces (like ESD) (ibid); and developmental evaluation as described by Patton (2011).

References for Section 3A

- Britt, H. (2016). Discussion Note: Complexity-Aware Monitoring. USAID, Washington, DC.
- Brouselle, A. and J. Buregeya (2018). Theory-based evaluations: Framing the existence of a new theory in evaluation and the rise of the 5th generation. *Evaluation*, 24(2), pp.153-168.
- Funnell, S. C. and R.J. Rogers (2011). *Purposeful Program Theory: Effective Use of Theories of Change and Logic Models*. Jossey-Bass/Wiley, New York.
- Guba, E.G. and Y.S. Lincoln (1998). *Fourth Generation Evaluation*. Sage, New York.
- Patton, M.Q. (2011). *Developmental Evaluation: Applying complexity concepts to enhance innovation and use*. The Guilford Press, New York.
- Rogers, P. J. (2008). Using Programme Theory to Evaluate Complicated and Complex Aspects of Interventions. *Evaluation*, 14(1), 29–48. <https://doi.org/10.1177/1356389007084674>
- Santos, BdS. (2018). *The End of the Cognitive Empire. The coming of age of epistemologies of the South*. Duke University Press, Durham and London.
- Stronach, I. and M. Mclure (1997). An optimistic deconstruction of research compromises in contract research and evaluation. In *Educational Research Undone: The postmodern embrace* Open University Press, London.
- Weiss, C. H. (1997). Theory-based evaluation: Past, present, and future. *New Directions for Evaluation*, 1997(76), pp. 41–55. <https://doi.org/10.1002/ev.1086>
- Wenger-Trayner, E. and B. Wenger-Trayner (2020). *Learning to Make a Difference: Value creation in social learning spaces*. Cambridge University Press, Cambridge.

Broad Approaches – Key Readings

- Key readings on the landscape of and different approaches to evaluation. These are meta-papers, rather than details about specific methods:
- Brouselle, A. and J. Bureyega (2018). Theory-based evaluations: Framing the existence of a new theory in evaluation and the rise of the 5th generation. *Evaluation*, Vol. 24(2), pp. 153–168.
- Funnell, S. C., & P.J. Rogers (2011). *Purposeful Program Theory: Effective Use of Theories of Change and Logic Models*. Jossey-Bass/Wiley, New York.
- Pawson, R. and N. Tilley (1997). A history of evaluation in 28½ pages. Chapter 1 in *Realistic Evaluation*. Sage, London, pp. 1-29.

3B FORMAL METHODS FOR EVALUATING ESD

3B.1 Empiricist Approaches

This section introduces two very commonly used methods for evaluation in ESD, namely experiments and surveys. We discuss them as examples of empiricist approaches to evaluation. As outlined in section 3A, these are approaches that privilege measurement, in the form of quantitative data that can be statistically manipulated, to the exclusion of all other forms of information. The two methods introduced here can be usefully combined with other approaches, but then the empiricist underpinnings - that only that which can be measured, actually exists - would have to be discarded. (In a critique of the empiricism of logical-positivist approaches to research, Arjen Wals argues that “What you can’t measure still exists” and can be extremely important for evaluating ESD).



3B.1.1 Experimental Method

Evaluations based on experimental designs, are often referred to as the ‘gold standard’, because from the empiricist (measurement-driven) approach, they are the ideal way to evaluate the effectiveness of educational and development programmes. Why is this the case, when this method has serious shortcomings?

In fields like physics, engineering, forestry and agriculture, it is possible to do experiments and draw strong conclusions from them. For example, farmers can experiment with farming methods: They can divide their maize field in two, and treat the one field with organic fertilizer, and the other with inorganic fertilizer and then measure, based on the tons of maize produced, which treatment (fertilizer) was the most effective for growing maize under the prevailing conditions. They can do so with some certainty if they are able to control all the variables that foreseeably influence maize yield: the same seed, soil, amount of water, temperature and other conditions, for the two experimental plots.

Can ESD evaluators do the same? Can we evaluate which of our ESD strategies are more effective for producing knowledge, agency, action competence, and other desired outcomes as outlined in Module 1?

This is the intention when evaluators design baseline tests or surveys, introduce an ESD programme or activity, and then repeat the test or survey afterwards to see if there is a change (in what is known as a pre-test, post-test or quasi- experimental design; if there is not a comparable group that do not get the programme, which would be a true experimental design).

Unlike the farmer, the educator does not have a relatively¹ closed system with relatively few variables that are reasonably easy to manipulate. Unlike maize plants from the same seed, human beings as learners vary tremendously, because they bring a variety of personal traits and social circumstances, educational and cultural backgrounds and life histories, into the ESD situation. So, some participants in an ESD programme may respond very positively to the activities on offer (the 'treatment') and others may respond very negatively to the same activities; think of Republicans and Democrats attending the same programme on climate change, for example.

For both of these groups there could be significant changes, but when we add it all up to measure the effectiveness of the programme as a whole, the positive and negative change cancel each other out and they might have to report, "no change" on average, which would obviously not be that helpful if we would like to know, why did some experience a positive change and others a negative change?

Furthermore, experimental designs are often unsuitable for evaluating the effectiveness or value of ESD programmes because there are numerous 'interfering' variables that make it difficult to attribute any observed change to the programme that is being evaluated; e.g., some participants may also be watching ESD related programmes on TV or YouTube, some may have parents who involve them in eco-friendly projects, and so on. Evaluators cannot eliminate these real-life influences of which there are many (we encourage you to think of others). The ESD educator is also part of the system; one lecturer, for example, will present the same ESD course differently from another lecturer, depending on their personality, preferences and nuanced differences in skills; and the same lecturer will do it differently from day to day, or year to year, depending on, say, how they relate to the particular group of students or what else they

¹As parts of modified ecosystems, a field of maize is of course not an entirely closed system, it is open to the weather, atmosphere, passing pathogens, insects and possibly covert maize harvesters, among other unpredictable variables.

have on the go at the time. The multitude of variables influencing how ESD activities are delivered and experienced, are typical of the complex open systems that make up educational settings. If one tries to reduce this complexity so as to do an experiment (such as a pre-test, post-test design) then it becomes questionable as to whether the conclusions from such a study, would actually apply in the real world, given that the latter is much 'messier' (see Complexity-sensitive approaches).

Another challenge with (quasi) experimental evaluation designs, is the extent to which one can precisely measure outcomes of ESD programmes, in the same way that one can weigh the amount of maize produced in an experimental field. What gives rise to this challenge? Think back to Modules 1 and 2 and answer this question yourself.

We can to some extent measure what people have learned in an ESD activity or programme, in terms of new facts and insights. But based on Modules 1 and 2, would knowledge be the sum total of what we want to see come out of ESD activities? How would one precisely measure the other intended outcomes? And, to what extent is it possible to predetermine intended ESD outcomes? In some situations, the nature of the desired outcomes may need to emerge from the programme, and then it is not possible to do a pre-test to determine a baseline of those related variables.

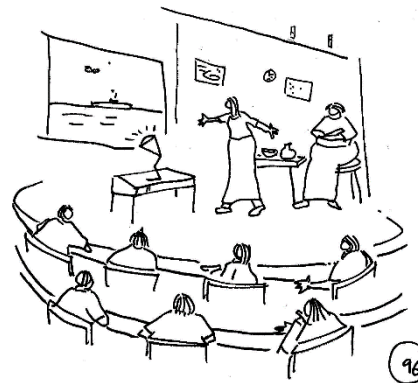
But the problem with using an experimental design to evaluate ESD is not just a practical one, which can foreseeably be addressed through more precise instruments to measure changes in knowledge or behaviors. There is also a problem with the assumption that ESD is a discrete intervention that brings about a discrete change in a particular space of time. Referring back to Modules 1 and 2 will suggest reasons why this assumption is flawed.

Does that then mean that experiments have no role in ESD evaluations? On the contrary! In some situations, experiments can be very useful in ESD programmes, not to evaluate the programmes, but to actually roll them out, in an evaluative manner.

ESD educators can work with farmers to help them test eco-friendly and climate-smart farming practices, by using experiments. If farmers have not used experiments before, the basic principles of controlling variables, carefully measuring, recording and then comparing results, should be explained, and this will improve farmers' adaptive capacity to mitigate climate change impacts like droughts, for example. The

strength of the experiment in simple systems where it can be applied, is the systematic approach to testing. When farmers can do such reliable experiments themselves, or in collectives, they are not dependent on commercial companies who may have only a short-term incentive to sell whatever products they have, rather than to attend to what farmers or the environment need. In an ideal scenario, farmers and agribusiness would work together to experiment with new (and sometimes old) varieties to make farming and food production more sustainable.

In a university situation, ESD lecturers could experiment with different ways in which to engage students in ESD content, and they could invite students to be part of the experimentation. A pre-course questionnaire could be used not only to establish what learners already know, but also to adjust the course content (e.g., perhaps putting more emphasis on what is not yet known, while building on what is already known). Instead of a comparative post-test questionnaire at the end of the course, students could be asked to give evaluative feedback throughout the course, which would help lecturers to adjust, but if well designed, the feedback activities could also help students to become reflexive about their own role and learning in the course, and, if they are education students, about course design as well.



Later in this module, the design of realist evaluations for evaluating what works for whom, under what circumstances, and why, is discussed. This is an example of how ESD evaluators can use the power of the scientific method, like experiments, in complex social systems, but with different assumptions to the empiricism that underpins the 'gold standard' approach.

References for Experimental Method

- Wals, A.E.J. 1993. What you can't measure still exists. In Mrazek, R. (Ed.), *Alternative paradigms in environmental education research*. Troy, Ohio: The North American Association for Environmental Education (NAEE), pp.297–298.

3B.1.2 Survey Method

Surveys are frequently used to gather data for evaluations, for example to canvas opinions on whether intended outcomes have been met, or to establish a baseline for later comparison. A description of this method will be added to future iterations of the course materials. In the meantime, many research methodology texts have good guidelines on the design of surveys, survey tools (questionnaires), analysis methods and software packages for large samples, and the interpretation of the data for evaluation purposes. An example is “Fundamentals of survey research methodology”, a guideline prepared by Glasgow, P.A. (2005) for Mitre, Washington C3 Centre, Mclean, Virginia
https://www.mitre.org/sites/default/files/pdf/05_0638.pdf).

3B.2 Constructivist Approaches to Evaluation

The next section introduces two evaluation methods that are not based on empiricist assumptions, but rather on constructivist assumptions. In constructivist approaches, human and social constructions, like opinions, perceptions, interpretations, cultural values and institutional norms, are also valuable data in helping evaluators understand what is going on, and whether it is worthwhile or not.

3B.2.1 Most Significant Change (Story) Method

Introduction

The Most Significant Change (MSC) evaluation method is both constructivist and participatory in nature. It also exemplifies a process-based approach, and is most often used for evaluative monitoring, but can also be used for evaluation. Since its development in a complex rural development programme in Bangladesh by Davies (1996), and further development with Dart and others, it has been used and modified widely around the world.

Other names for it include non-indicator-based M&E, the ‘story approach’, impact monitoring, and evolutionary approach to organisational learning. This summary is based on a comprehensive guide by the main proponents of the method, responsible for its initial development, Davies and Dart (2005).

How it Works

In essence, the method involves sourcing, selecting and learning from what a hierarchy of stakeholder's regard as the most significant changes that have taken place as a result of an intervention. Field staff source stories from intended programme beneficiaries, who then get together to select the top stories from among them, to share with seniors in the programme or organisations. These stories are not fictions, but narrative, experience-based accounts of what the beneficiaries, on careful reflection, regard as the most significant change that has come about through the programme so far. Some evaluators prefer to call them accounts, rather than stories, if there is a risk that participants will think they need to 'make up' an illustration of what success would look like. The narrative accounts take the form of what was done or happened, who was involved/who did it, why was it done/did it happen, and why is it significant. They may be shared verbally, but they are, importantly, also written up.



Figure 2: The questions structuring the most significant change story / account (Davies and Dart, 2005)

Once each participant has provided their story, it is discussed among fellow beneficiaries, and field staff as a collective. The discussion is aimed at identifying which of these stories are the most important, and results in choosing an agreed number, say seven. The criteria used for choosing, are articulated and recorded. The selection and the reasons for the selection are then shared 'upwards', e.g., to programme managers.

The managers then repeat the selection process, by choosing even fewer stories (say three of the initial seven) that they regard, after due deliberation, as representing most significant changes that have come about in the life of this programme. The managers in turn share their selection and the reasons for the selection with the top structure, e.g., the funders or government department within which the organisation or programme functions. This top structure then repeats the process to come up with just one story (out of the three they received) that reflects the most significant change the programme has brought about.

An important aspect of the process is to communicate back to the lower levels in the hierarchy, what the next level up has decided, i.e., what their selection was, and why. The reactions to the next layers' selection can be discussed, and the gist of this discussion can also be fed back up to the next layer, creating an interactive dialogic process.

The MSC method is not prescriptive, but there are a few key principles and guidelines. For example, it should be used repeatedly over the life of the programme being evaluated; the stories should be verified; and the change domains and reporting period should be defined. Secondary analysis (from time to time) and meta-monitoring (ongoing) add rigor and depth to the MSC method. For how to do this, how to undertake the analyses, and other practical hints, see "Overview of implementation steps" (Davies and Dart, 2005, p.10) and Chapter 2 of that Guideline.

ESD example:

Say a funder teams up with a national department of higher education, to start a comprehensive ESD programme in universities. Each university receives a grant and an open remit to promote environmental sustainability at their institution over a five-year period. After a year, ESD lecturers collect stories from senior students as to the MSC they have seen at the university this year, with regards to sustainability. They deliberate these accounts with the students and each other, and select the top 5 stories that best represent the kinds of changes they regard as significant (negative or positive).

1. a brand new e-waste reduction project started
2. revival of a dormant paper recycling project
3. introduction of ESD in the form of a lecture during the first years' induction week
4. re-routing of a portion of research funding to ESD topics

5. introduction of a prize for best ESD champion.

Along with the reasons for these choices (why they value these) this selection is then shared with the university senior management. Senior management then choose 2 of these stories as the most important, and communicate these up to the programme steering committee consisting of funder and government representatives. Say in this case, they do not support prizes for champions that much, because they believe ESD should be a collective effort and have intrinsic motivation, so #5 would not be selected. They also don't think a once off lecture will make that much of a difference, so #3 is also not selected.

The steering committee now receives stories 1, 2 and 4. After much deliberation they choose the new e-waste programme, with revival of a dormant recycling programme being a close second, because they also think that sustainability of initiatives are important. They like the e-waste initiative because it is cutting edge, is a show-stopper example of innovation, has potential for research opportunities and further innovation, and e-waste is a particularly big new concern for universities, that they should be best placed to address (as centres of innovation). They do worry that there is not a strong curriculum change story yet, and indicate that they look forward to this coming up in the next round, at the end of year 2.

Communicating all these decisions up and down the hierarchy helps all the role players understand what is collectively valued. Lecturers might well start on a more comprehensive curriculum change programme, perhaps linked to innovations like e-waste, and senior management might also take note that students value the extra push that comes with a financial incentive and recognition.

Value of the MSC Story Method

The value of this deliberative and participatory approach includes:

- it monitors (intermediate) outcomes and impacts, rather than just activities and outputs
- many stakeholders get to share what they regard as important, what change is desirable, and what is happening in the programme that should be noted by all

- being explicitly open-ended, it picks up changes that could not have been foreseen beforehand, therefore overcoming the limits of M&E where only pre-specified indicators are being monitored, and where unintended consequences and impacts might be missed
- it allows for progressive focusing and consensus building, which is valuable when there is disagreement or lack of clarity as to what a programme is or should be doing
- the process of coming up with stories and selecting among a diversity of stories, focusses all participants' minds towards impact, and can therefore steer programme activities in that direction
- it can prevent misdirected activities and mission drift if stakeholders hold each other accountable; if the selected stories are very far off from the initial starting points of the programme, then stakeholders have an opportunity to discuss whether this is a good thing or not, i.e., were the initial starting points inappropriate, or is there an inappropriate deviation from what stakeholders' value
- therefore, the method combines open endedness and flexibility well with focus
- combining process-based, participatory, constructivist, dialogic approaches
- programme sponsors can decide, based on what accounts are available and selected, whether this is the kind of programme (and organisation) that they want to continue investing in, because the available stories present what is actually happening, and the selections present what the participants value
- providing volumes of text-based material that can be used further for evaluation processes and for communication purposes; evaluation reports with actual stories of change are ideal for sharing what a programme is about, with a range of audiences
- the method fosters organisational learning at multiple levels, but also accountability - change stories must be verified, and if there are no or few significant change stories to share, it indicates that the programme is not doing well.

Underlying theory

The constructivist theory of evaluation implies that what people value, or regard as important, is an important source of information or knowledge in evaluation (Guba and Lincoln, 1989). To know what people value and whether they think that some value has or has not been achieved, is a key knowledge consideration in programmes that are designed to benefit people. However, participants may knowingly or unknowingly give a narrative account of a change ('construct' a change) which has not actually happened. For example, I

might say that the top management of my university has adopted ESD, when there is no other evidence for this, or even contradictory evidence. In the MSC method, the stories provided must be verified, by other participants and/or other forms of evidence (e.g., in this case, an ESD policy document endorsed by the top management, or a significant budget allocation for ESD at my university). MSC insists that not all of reality is socially constructed (i.e., whether a budget has been allocated or not, is not a construction) and at the same time, that people's perceptions (e.g., whether the budget was a significant shift to ESD or not) comprise valuable information. That is, the MSC uses a constructivist epistemology, and a realist ontology.

The MSC is also underpinned by organisational learning theory. Davies drew on a learning theory called evolutionary epistemology. Learning is seen as the selective retention of information. The evolutionary learning process is seen to involve the reiteration of variation, selection and retention processes. The sourcing of accounts of change from across programmes provide the variation; the MSC process of choosing the stories regarded as most important involves the selection, and the adoption of one story above others, encourages the retention of the knowledge of what is happening and valued in a programme. In this regard it also draws on Bateson's first and second order learning (in the first order we learn what is happening and what we and others value; and in the second order we learn that we might need to revise what we have regarded as important, because there are new and even more important considerations that shape what we value - this is often framed as a revision of fundamental goals). The organisation (programme) becomes more conscious about what it values (the selected story presents the 'best fit' with organisational values) and opportunities are opened up for changing practices to better achieve more of what is valued, as represented by the chosen story or stories.

Limitations and complementary methods

The MSC method has some limitations. Verifying stories of change can be very difficult, for example. The process of selecting and celebrating one story only, could fail to portray or affirm a variety of valuable processes in complex programmes. The MSC method can only be used in programmes where social change is desirable, and in some programmes, consolidation or conservation rather than change, might be the desired outcome. Another limitation relates to the design which requires hierarchical structures. Although many organisations and programmes function well as hierarchies, these are limiting if there are no good opportunities for dialogic communication and mutual respect across the levels in the programme

or organisation. Many programmes consist of multiple organisations networked together, rather than clear cut hierarchical relationships to each other. For this reason, the MSC method has been adapted to multi-partner, non-hierarchical situations.

Other limitations can be addressed by complementing the MSC method with other methods. For example, it can be complemented with biophysical monitoring, with theory of change development, and with the use of predetermined indicators based on logic models.

References for Most Significant Change (Story) Method

- Davies, R. & Dart, J. 2005. The 'Most Significant Change' (MSDC) Technique. A guide to its use. www.mande.co.uk/docs/MSCGuide.htm and www.clearhorizon.com.au

3B.2.2 Appreciative Enquiry Method

What is it about?

Appreciative enquiry is an approach to evaluation (Preskill & Catsambas, 2006, Preskill & Coghlan, 2004), that is also an approach to organisational development (Cooperrider et al., 1995; Cooperrider, Whitney, Stravos, 2002). Appreciative enquiry has similarities with outcomes harvesting, and the Most Significant Change method, in that what is to be observed by the evaluation, is not tightly specified beforehand.

As the name suggests, appreciative enquiry chooses to focus on identifying strengths and then developing plans to build on these strengths, where other evaluation and organisational development approaches might focus on identifying problems and gaps.

The process is often structured (according to Funnel and Rogers, 2011) around what is known as the 4-D model consisting of:

1. **Discovering** - when the organisation is at its best. This is identified through people in the organisation talking to each other, often in structured interviews, and describing these situations as richly as possible.

2. **Dreaming** - this phase is often run in a large group where organisation members come together in a 'conference' type session where they are encouraged to picture (envision) the organisation as if the peak moments identified in the discovery phase, are the norm, rather than the exception.
3. **Designing** - a small team is then mandated to go away and design ways of creating the organisation as designed in the big group conference, and finally,
4. **Destiny** - is the phase where the changes are to be implemented.

The appreciative enquiry method has also been adapted to community capacity development, using a Seven D method, in which the participants work systematically through seven iterative cycles involving the following seven components:

1. Developing relationships - to assist the community in their endeavors
2. Discovering the community's strengths and needs
3. Dreaming the future they want to see
4. Directions - strengthening their ability to analyse their potential and challenges, to reach consensual decisions on collective action - leading into the next step -
5. Designing the actions to take
6. Delivery - taking these actions, perhaps with partners
7. Documentation - this includes recording the process and decisions made as well as reflecting on the outcomes (see Dhamoratan, 2009)

When is it used and what is its value?

An appreciative enquiry can be used in the early stages of a programme or organisation's life, and during times when there might be quite a few problems and a need to lift morale and energy to build on what strengths there are. It is based on the assumption that which we focus on, will grow, so that it is better not to focus on the problems in a time when there is a strong need to build on and expand the strengths. It assumes that the problems are already well known. Appreciative enquiries are developmental in nature, that is, they help to shape and (re)direct programmes and organisations in desirable directions. Many aspects of the 4-D and 7-D processes are participatory in nature. They would ideally be used iteratively throughout the life of a programme or organisation, for ongoing learning.

ESD Example

A fictional ESD example could be a vocational college that is struggling to bring ESD into its curriculum, due to a number of systemic factors that almost seem insurmountable to overcome. Morale among the lecturers who believe that ESD is very important, is low and they feel their problems are so many that they do not know where to start. They are unclear on how exactly to bring about curriculum change, they do have clear ideas for greening their campus but they lack the financial resources to buy the green technology they want, like water tanks and solar panels, and they have little support from top management who prioritize marketing of the college and does not think that ESD is a priority. An appreciative enquiry evaluation works with the lecturers to focus on what they do have: enthusiastic students who are active on social media; clear ideas for greening the campus; and good partnerships with a local development agency. Building on these strengths, they design a plan to raise funds for the green technology through their existing partnership; to install the water tanks and solar panels, invite senior management to launch the new facilities, and give the event exposure through their enthusiastic students' media campaign; then to build on this success by inviting curriculum experts to assist them in a curriculum design process around the new technology, and once again, inviting senior management to participate in the activity and demonstrate its value for marketing the college as responding to societal needs and demonstrating new technology through ESD. Rob O'Donoghue used an appreciative enquiry method to evaluate the Regional Centre of Excellence (RCE) in Makhana (see case study, Module 2).

What are the limitations and how can these be overcome?

Appreciative enquiries will not provide insights into the nature of the problem being addressed, and from time to time (including right at the start of a programme) it would be necessary to try and unpack the causes and consequences of the issue(s) being addressed. Therefore, appreciative enquiry is well complemented by problem tree analyses, and/or a situational or contextual driver analysis, and similar processes. These might be indicated when, for example, the planned design interventions do not yield the intended outcomes. For example, if the lecturers in our example are not able to raise funding through their partnership, it could be helpful to analyse why this is the case: do they lack budgeting skills, are they focusing on the wrong funding sources,



or is the current climate such that donors are unlikely to fund colleges, and other greening avenues that do not require donor funding, would be more likely to succeed?

Appreciative enquiry is also well complemented (followed up or interspersed) with evaluations that use indicators to look for specific outcomes, based on the earlier appreciative enquiry phase: Have we raised the money? (If not, why not?) Have we put up the water tanks and solar panels? (If not, why not?) Have we engaged our students and senior management effectively in the process? (If not, why not?) Have we done something else instead and has that also/better met our intention, of introducing ESD at the college?

References for Appreciative Enquiry Method

- Cooperrider, D.L., Barrett, F. and S. Srivastva (1995). Social construction and appreciative enquiry: A journey in organisational theory. In Hosking, D., Dachler, P. & Gergen, K. (eds). *Management and Organisation: Relational alternatives to individualism*. Avebury Press, Aldershot.
- Cooperrider, D.L., Whitney, D. and J.M. Stavros (2002). *Appreciative Inquiry: The handbook*. Crown Custom Publishers, Brunswick.
- Preskill, H. and T. Catsambas (2006). *Reframing Evaluation through Appreciative Inquiry*. Thousand Oaks/Sage, California.
- Preskill, H. and A.T. Coghlan (Eds) (2004). *Using Appreciative Inquiry in Evaluation*. New Directions for Evaluation, 100. Jossey-Bass, San Francisco.

3B.3 Process-based evaluations

3B.3.1 Value-creation evaluation

What is it about?

Learning processes like ESD create different types of value that each contribute, singly but often in combination, to the achievement of desired outcomes. This method aims to evaluate social learning processes by identifying these different kinds of value through the use of a practice-based framework or template, while at the same time (as the value-creation stories are elicited and shared) also fostering such social learning processes.

The method was originated by Wenger, Trayner and De Laat (2011) for evaluation of learning in networks and communities of practice, and has been applied in and later extended to broader social learning spaces (Wenger-Trayner and Wenger-Trayner, 2020).

The value-creation framework identifies and describes the following types of value that arises from social learning:

- **immediate value** - not necessarily immediate in terms of time, but directly emerging from the social learning process being evaluated
- **potential value** - what participants take away with them from the social learning process, that could be used to further create value
- **applied value** - arising from using what was potentially useful, or arising from application of learning in practical change-oriented activities
- **realised value** - aspects of the desired change are being achieved
- **reframing value** - the problem being addressed, or the way in which to achieve the desired change, is being differently (better) understood and framed in such a way that it can be addressed with greater success.

The above each constitute a cycle of learning and value creation, and are seldom instantly achieved. They are interactive, and more than just single steps. Additional value is created, towards the achievement of the desired change, through a flow from one type of value creation (cycle) to another. There can also be feedback loops back to earlier cycles, with additional value generation then further becoming possible. Furthermore, the framework also involves orienting, enabling, strategic and framing value. This will be illustrated in the ESD example.

ESD Example

In this fictional example, university lecturers from South Africa, Mexico, India and Germany come together in Mexico with some expert mentors, to exchange experiences. Their aim is to learn from each other and the expert mentors, how best to introduce ESD at their respective workplaces. Thereafter they return home and try to introduce ESD at their institutions. Six months later they are interviewed to establish what value, if any, had been generated through the learning exchange.

Some participants narrate how they related very easily to the others on the exchange, and found the experts to be both friendly and knowledgeable. They felt that the exchange was highly relevant to their aims, well planned and very energizing (immediate value). They left the event feeling inspired, with lots of new knowledge and new contacts (potential value). Most followed up on these contacts and they kept touch with each other, which led to further inspiration and more information, on how to practically use the knowledge to set up new ESD programmes (applied value). Some learning participants used all the preceding value examples, to start new ESD programmes (realized value) - their aims were achieved.

They also realized that they needed to bring their university structures on board, and drew on the experts to help them motivate and bring on board the universities' senior management; this is an example of strategic value, because it will allow them to expand and continue these programmes (e.g. introduce them to other departments and campus management).

As the ESD lecturers worked with their new programmes, they became quite knowledgeable themselves and started revising what they had learnt on the exchange, and also feeding that back into the original network (both reframing and framing value).

In the evaluation report there was also a record of the orienting value of the many years of experience of the experts who joined and hosted the event, and the enabling value created by the funder who made the exchange possible in the first place.

How it works

There are a number of different ways in which facilitators and participants can be convened to reflect on the social learning process(es) being evaluated, and decide what value was generated, positive or negative, or not generated. Evaluators (who are often also learning facilitators) work with and alongside participants in the social learning process that is to be evaluated. Participants also do the evaluation themselves, if they are familiar with the framework.

Value-creation 'stories' (not fictions, but experience-based accounts of what actually happened) are elicited from participants, with or without using the given framework. These narratives are then analysed to identify instances of value created, which are plotted on to the framework, with associated evidence. The framework gives possible examples of each of the value types, whether positive or negative. For instance, participants would be prompted to ask if any knowledge has been introduced into the social learning space being evaluated. Participants could have examples of positive value being created with regards to knowledge, e.g., relevant information was well presented and inspired them to make changes in their workplace. A negative value created could be that there was too much new knowledge which overwhelmed them so that they left the programme; or relevant knowledge was poorly presented and left them unclear on how to continue. An example of zero value created could be that there was no relevant information available at all. The facilitator would log "knowledge" as an example of "potential value" that was either a positive value created, or a negative value created, or no value created at all.

Cycles of value created, flows of value across the various cycles, and loops of value-creation back into the value chain, or even value branches into adjacent social learning spaces, can be identified through further analysis. This presents a picture, either simple or elaborate, of different kinds of value created, which can then be shared with sponsors, facilitators, participants and other stakeholders, to inform further social learning processes.

Value of the method

The value-creation method is specifically designed to evaluate processes of learning, often throughout extended periods, and so it addresses the fact that the outcomes of change-oriented social learning processes are not only not always immediate, their benefits often only materialize in the long-term. Desirable outcomes of learning processes are also not always easy to tightly specify beforehand. The framework allows for unexpected positive value as well as unexpected negative value created by the process, to be reported. Hence this is a useful method to use when process-oriented and complexity-sensitive evaluations are required. It allows for high degrees of uncertainty, while at the same time providing a clear picture of value that has emerged in the process (thus far).

This picture of value created can be visually presented using the given framework of cycles, flows and loops, and therefore it can also be quite easily communicated visually, with accompanying text, and complementary examples of value-creation stories. It provides for a good blend of qualitative and semi-quantified findings.

This is also a participatory evaluation process, in which what counts as 'value' or meaningful change, is deliberated among participants. However, it is not a case of random subjectivity or 'anything goes', as decisions about and evidence for value created, are based on clear guidelines and indicators which have also been deliberated among participants.

Underlying theory

The value-creation method is informed by a range of fields, from business studies and economics, which originated the idea of a value chain (Potter); to the roots of international development theory, in political economy, anthropology, development theory and social theory. Philosophically, it has humanistic roots and also aligns most closely with an interpretivist paradigm and constructivism (Wenger-Trayner and Wenger-Trayner, 2020, p.53). From social theory it draws on activity theory, and from development theory, it draws on the capability's theory of Amartya Sen, Martha Nussbaum and Robeyns.

Limitations and complementary methods

One limitation of the value-creation method is ironically that it can be readily understood; to any facilitator of learning processes, the categories of immediate and potential value, applied and realised value readily make sense. The danger is that they can be interpreted in a somewhat superficial, everyday sense of the words, whereas there is a sophisticated theory and methodology behind how the categories should be defined, interpreted and used. This limitation can be overcome by working closely with the many accessible descriptions of the method, the most comprehensive possibly being Wenger-Trayner and Wenger-Trayner (2020).

References for Value Creation Framework

- Wenger, E., Trayner, B. and M. De Laat (2011). Promoting and assessing value creation in communities and networks: a conceptual framework. Rapport 18, Ruud de Moor Centrum, Open University of the Netherlands. <https://wenger-trayner.com/resources/publications/evaluation-framework/>
- Wenger-Trayner, E. and B. Wenger-Trayner (2020). Learning to Make a Difference: Value creation in social learning spaces. Cambridge University Press, Cambridge.
- Wenger-Trayner, E., Fenton-O'Creevy, M., Hutchinson, S., Kubiak, C. and B. Wenger-Trayner (2015). Learning in Landscapes of Practice: Boundaries, identity, and knowledgeability in practice-based learning. Routledge, London and New York.

3B.3.2 Developmental Evaluation See Complexity-sensitive Approaches.

3B.4 Theory-driven Approaches

As introduced in Module 3A, theory-driven approaches are aimed at evaluating societal and institutional intervention programmes, such as the introduction of a new ESD policy for schools, or an intervention to better resource ESD at all universities, or the implementation of a programme to strengthen farmers' ability to use new sustainable agriculture practices.

Surely all programme designers have on theories of how the change they want to see in the world, will come about through this programme? That is indeed so, but in practice, few are explicit about what exactly this theory is, which makes it difficult to evaluate the programme, unless of course, the evaluation is actually aimed at identifying the theory. In theory-driven approaches to evaluation one may therefore start by surfacing and articulating the programme theory, then designing the evaluation from there. Or, one can intentionally undertake an evaluation (often a meta-evaluation) to develop theory.

The interest in the theoretical underpinnings of interventions is not new, although it is enjoying a particularly popular moment at this time. In their very practical handbook, *Purposeful Programme Theory: Effective use of theories of change and logic models*, Funnel and Rogers (2011) traced the history of theory-driven approaches all the way back to Kirkpatrick who, in the 1960s, used four domains of change for evaluating training.

In more recent times theory-driven evaluations have become popular where programme developers recognized complexity in the situations they were trying to change, to the extent that there is now talk of it representing a fifth generation of evaluation theory (Brouselle & Buregeya, 2018).

3B.4.1 Programme evaluation with theories of change

A programme theory is an explicit theory or model of how a change will occur and how an intervention will produce these causal processes (Funnel and Rogers, 2011, p.13) or put in programme language, how an intervention contributes to a set of specific outcomes through a series of intermediate results (ibid, p.31).

Programme theory consists of two connected components, the theory about how the desired change will come about, and the implementation or action theory (Weiss, 1997) which explains how the programme or other intervention is constructed to activate the theory of change; i.e., it explains the implementation activities. Other terms for programme theory are: theories of change or change theory, programme logic, and intervention logic. The ToC is part of the programme theory, but refers specifically to the central mechanism by which change comes about, for individuals, groups and communities.

As you can imagine, it is not always easy to precisely and comprehensively describe the desired change and how it is going to come about, and some evaluations are in fact aimed at developing the theory. When programme theories are used to design programme evaluations, this is known specifically as a theory-driven evaluation.

Our theory-driven evaluation method example is on how to construct a theory of change (ToC) in one programme, and how to then design an evaluation for that programme, based on the ToC.

Value of the Programme Theory Method

Programme theory methods have gained popularity in the past 30 years. They can be used for programme or intervention design, implementation, evaluation and ongoing learning and adaptation or revision of the programme.

The benefits include being able to design an evaluation that focuses on what the programme designers, sponsors, implementers and other stakeholders really want to focus on, as the crux of the change theory, as opposed to what is easy or conventional to evaluate.

By involving programme funders, designers, implementers and even beneficiaries² in articulating the programme theory (what is the change we want to see, how is it going to happen, and how is our programme going to make that happen) the method can help to develop agreement among the stakeholders about they

²Hanli Human conducted an evaluation research study in which she involved stakeholders in helping to identify programme indicators and theories of change, for the Tsitsa Programme (<https://www.ru.ac.za/elrc/projects/thetsitsaproject/>)

are trying to do, why and how; it can also help to improve plans by highlighting gaps and opportunities for collaboration; it can help to set realistic objectives; and guide the choice of meaningful indicators to track progress and identify successes. Importantly, in the case of realist approaches to programme theory, the method can also identify where and why programmes are failing; similarly it goes beyond evaluating whether programmes have worked as planned, to identifying how and why they have worked (identifying the underlying mechanisms of change that were triggered by the programme - what exactly it is that makes the programme work) so that it can then be introduced elsewhere as well, with some confidence that the underlying mechanisms will again be triggered, even in a different context.

Programme theory unpacks the closed box that contains the mechanisms that makes programmes work, to the extent that they do, for some people and in some circumstances. This allows us to further build education and development theory and inform future programme designs.

How the Method is Used

Say we want to evaluate a national programme aimed at encouraging more school principals to introduce ESD in their schools. There is funding for one workshop in each district, so principals will have an opportunity to attend one workshop. In the workshop the inputs are lectures and resources by ESD experts; talks by education officials; and opportunity to engage informally with other principals. Nine months after the workshops, evaluators ask the participating principals to share any examples of how they have implemented a whole-school approach to ESD in their schools.

Any number of outcomes are possible:

- A. Fewer than 10% of principals have introduced ESD - the programme has failed
- B. Some 90% of principals introduced ESD at their schools - a programme success!
- C. About half of the principals introduced ESD and the other half did not - a mixed result

In order to decide if the programme should be repeated the next year, or whether changes are needed, the evaluation should really explain why the particular results were achieved. In the case of a failure (A) a redesign is needed, but it would help to know why the programme failed. In the case of a resounding success (B), if the programme is to be rolled out in other countries, then we need to know what it was about

the programme that made it work so well in this context, so that the same elements can be retained in the out scaling or upscaling of the programme into other contexts. For example, if the underlying mechanism was that the programme provided knowledge that principals did not have before, then the programme might not be as successful in other contexts where school principals already have a lot of knowledge about ESD. On the other hand, if the programme was a success because the principals had an opportunity to meet with fellow educational leaders, with whom they kept in touch afterwards to share ideas on how to make ESD work in their schools, then it would be important not to cut the opportunity to interact from the workshop programme - which one might be tempted to do, if one assumes that the change comes about because a lack of knowledge is being addressed, as the basis for the change.

In the case of the mixed result (C) further investigation might show that the principals who did not introduce ESD were quite motivated to do so, but due to lack of internet connectivity at these schools they could not keep the connection with each other or find additional implementation resources online, while those principals who did introduce ESD, did.

Through in-depth interviews with principals and observations at schools, the evaluators could construct a theory of change looking like this:

access to new knowledge + motivation to meet national guidelines + collegial connections + aspiration to keep up with trends + access to implementation resources ⇒ introduction of ESD at schools under leadership of principals

In practice, the various elements of this 'outcomes logic' would interact with each other; for example, as ESD is introduced at the school, this may lead to new knowledge, more resources (e.g. attracting funding), further aspirations as praise starts to follow, and attracting more colleagues who want to learn from this initiative. Therefore, the process is often not linear but circular, or mutually reinforcing (and negative outcomes may also be based on feedbacks).

Realist programme evaluations were first developed by Pawson and Tilley (1997) and encourages evaluators to look for the underlying mechanisms that 'fire' in successful programmes. What was it about the workshop-based programmes that worked? Was it the information shared, the speeches, or the

formation of networks? If the speech from the government officials turns out to have been a significant feature of the programme, what was it about the speech? What did it elicit among the participants? Fear of not following policy? Inspiration? Aspiration to be a stand-out principal? Ambition to do well and become an official oneself? If the opportunity to network was the most significant driver of change for others, what was the underlying mechanism? Again, ambition to be a leader among leaders? Inspiration that social good can be done when working together, i.e., solidarity?

When many such evaluations are done, one may eventually identify mechanisms that drive change among school principals in particular contexts, and this knowledge - theory - can be used to design and plan new programmes, including ones that are not based on workshops, but still fire the underlying mechanisms. For example, if it appears that ambition is a driving factor, university certificate courses combined with national awards could be a powerful programme. If a desire for social solidarity is a driving factor, a course-activating learning network and shared change projects could be very powerful.

Some years later, an Education Faculty at a university then draws on this theory to design a course-activated ESD learning network for school principals. They are now in a position to explicitly articulate that they want to build a sense of community among the participants, and they design an evaluation that includes indicators of solidarity into the evaluation.

How to develop a theory of change

There are broadly three ways in which to develop a programme theory and the theory of change component of it. These are:

1. articulating the programme stakeholders' mental models
2. deducing the ToC from programme documentation, and
3. inductive reasoning based on observations of programme implementation.

In practice, one often uses a combination of these three.

Articulating the stakeholders' mental models involves interviewing them or bringing them together in focus group discussions or workshops. There are a number of ways in which stakeholders can surface and

articulate their mental models of the programme and why it will bring about the desired change. Keystone IPAL Guides 1 and 2 (2009) are among the many resources (another being Funnel and Rogers, 2011) that describe a good process for developing a ToC with stakeholders or programme participants. These processes often start, if there is enough time, with a problem analysis or a reminder of the underlying causes and consequences of the problem(s) that are being addressed by the programme. It also needs to include a consideration of other agencies involved in addressing this issue, and programme partners, and what they do. It is very helpful to bound the theory of change, that is, to indicate what is relevant but NOT being addressed by this programme.

For illustration purposes, a series of questions for drawing out theories of change, based on Funnel and Rogers (ibid) could be:

- Can you give us an example of where this programme is working really well? Why did you choose that example? What do you think is making it work well? (the why or underlying mechanisms)
- How would life be better for participants or intended beneficiaries of this programme worked well?
- What are the current barriers to a good life for programme participants, and how do you see this programme overcoming those barriers? What is it about how the programme operates that would or could make it produce the desired results or impacts?
- What else needs to happen? Who else needs to be involved?

The ToC needs to focus on explaining how the programme's activities contribute to the results, not just a list of activities followed by results.

How is the theory of change then used to design an evaluation?

1. Identify all the stakeholders in the evaluation
2. Identify how these stakeholders see the purpose and scope of the evaluation, and agree on a realistic purpose and scope (that is, to not try to do too much)
3. Translate the purpose in one or more evaluation questions
4. Identify what data is needed to answer the questions
5. Identify the data sources and methods that can be used to obtain the data

References for Programme evaluation with theories of change

- Funnel, S.C. and P.J. Rogers (2011). Chapter 2: Variations of program theory over time, pp. 27-30. In: Purposeful Program Theory. Effective use of theories of change and logic models. Jossey-Bass, Wiley, New York.
- Keystone IPAL Guide 1. (Undated) Developing a Theory of Change: A Framework for Accountability and Learning for Social Change. www.keystonereporting.org
- Keystone IPAL Guide 2. 2009. Developing a theory of change: A guide to developing a theory of change as a framework for inclusive dialogue, learning and accountability for social impact. www.KeystoneAccountability.org

3B.4.2 Realist meta-evaluation/synthesis

What is it about?

Realist synthesis or meta-evaluation is a systematic evidence review approach for drawing conclusions about what works, for whom and why, in complex social intervention. That is, they provide an explanatory analysis of the role of contextual factors and ‘firing mechanisms’ to the success or failure of an intervention. At the heart of it is the “programme theory”, that is an assumption of how a programme is expected to work. Evidence is systematically gathered to test and refine the programme theory (Pawson, et al., 2004).

At its core, a meta-review strives to shed light on a topic that is characterized by diverse philosophies and that is explored from varying angles. It does this by showing areas of divergence and convergence in how researchers explored the topic or related topics (Wong, et al., 2013). As such, in this design, the research team reviews hundreds or at least dozens of studies that hypothesize about how a desired social change comes about, and then construct a meta-level theory about it. In realist meta-evaluations (also known as realist synthesis) the proposed theory is in the form of: what works, for whom, why and under what conditions? These questions emerge from the understanding that the same programme may be replicated in multiple contexts and will produce varying outcomes as a result of contextual variations (Pawson, et al., 2004).

Thus, the implementation of a programme in multiple contexts with strict adherence to the terms of reference does not guarantee identical outcomes. However, this does not then restrict researchers to the confines of

one intervention or context with the same geographical location, culture, or language without a good reason. Alderson, Green, and Higgins (2004) note that restrictions can only be done for a good reason. As asserted by Westhorp (2014), it is therefore important that evaluators have an in-depth understanding of how the chosen interventions work for different groups and the outcome (or impact) data is disaggregated according to different groups (e.g. men, women, youth, older people, different castes).

When is it used and what is its value?

Realist evaluation is primarily concerned about explaining the reasons behind the performance or non-performance of programmes in different contexts. It is strongly interested in the causal relationships among project components and understanding how they work together to produce results (Westhorp, 2014).

Stern (2015) provides conditions under which realist evaluation would be an approach of choice:

- when evaluating new initiatives, pilot or innovative projects or programmes, or any other intervention where there is evidence that the project or programme works, but it is not yet understood how, why or for whom.
- when evaluating development interventions that are intended to be expanded, replicated, scaled-up or mainstreamed (in effect, any intervention that may need to be adapted to new contexts); and
- when evaluating projects or programmes that have previously had mixed results, in order to better understand why results have been inconsistent.

By asking the questions, “*what works for whom, in what contexts, in what respects and how*” realist evaluation takes the evaluation process a step further by recognizing the variabilities in personal, and socioeconomic contexts. It explores the subjective dispositions that influence how people interact with the mechanisms of an intervention at a personal and at a group level, which ultimately determine programme outcomes. These insights are useful for stakeholders in developing targeted interventions and in redirecting existing interventions. However, the insights are not quick to come by, they are products of iterative and cumulative processes of learning that go beyond a single intervention to informing other programmes.

ESD example

The following hypothetical ESD example of how to adopt a realist synthesis/meta evaluation is guided by Pawson et al. (2004) and Rycroft-Malone et al. (2013). The example will detail all the steps to be taken and the justification for each decision for potential users to replicate the process. It is imperative to mention that there is no one size fits all approach to realist reviews nor a dictated set of arbitrary standards (Alderson, Green & Higgins, 2004). The following stages only show the principles that should guide chosen approaches. Rycroft-Malone et al. (2013) recommends that whatever the choice the reviewer prefers, they should remain sympathetic to the philosophy of realism.

- **Clarify the scope of review**

Like all other review approaches, realist synthesis starts with conceptual framing and focusing and production of the research protocol. This involves understanding the phenomenon under study and the desired outcomes and having a clear and in this case detailed outline of how the review will be undertaken. Thereafter, questions that would guide the process to the desired outcome will be deliberated on and approved by stakeholders. This is the most painstaking and contentious part of the review, where evaluators and stakeholders have to go through an iterative process of focusing questions.

In a hypothetical example of the state of Climate Change Education (CCE) in tertiary institutions, stakeholders come together to discuss how CCE may be mainstreamed across all disciplines in all universities. A meta-review would involve determining “what works” based on what other evaluators have found and published in evaluation reports, which then become the data sources of the meta-evaluation or synthesis of the available knowledge. To commence the review, stakeholders will develop operational clarity on what counts as CCE and tertiary institutions and how wide the geographic scope of studies should be; also, what kinds of studies would be regarded as credible enough to include in the review. Pawson et al (2004) advise that all the decisions that are made at this stage should be detailed in a study protocol that shows the review process from the conception stage to finalization. The protocol kickstarts the programme, and is a publication in its own right.

- **Refine purpose of review**

This stage involves the articulation of the various themes that will guide the review. Pawson, et al (2004) outline four “cuts” of this stage, reviewing for programme theory integrity, reviewing to adjudicate between rival programme theories, reviewing the same theory in comparative settings, reviewing official expectations

against actual practice. These divisions are not independent of each other, in fact they inform one another and, in some instances, they are original versions and upgrades of each other. Following these divisions, in our hypothetical example of CCE, the first division would involve going through the history of CCE and identifying the misgivings of the past interventions and the challenges that caused them. The second division, relying on evidence, will involve further exploration and adjudication of the various approaches implemented in CCE, understanding how they work, and the motivations behind them. This stage is what drives realist reviews and sets it apart from other reviews. The third cut involves uncovering many studies of the CCE interventions in different contexts and mapping out patterns of the success and or failures of the interventions. This is done cognizant of the varying contexts and how they influenced the outcome with a special focus on the intended beneficiaries and how they relate to the programme indicators. The fourth cut is about the application of the aspects of the second division. This stage is characterised by various contentions that may include disagreements between policy and practice-oriented stakeholders on the best way to implement the programme. As was the case in the second division the contestations between theories is also part of this stage, these will be resolved by empirical adjudication. The consensus will be achieved through agreeing on the existing intervention/policy/theory to be used as the benchmark.

For this stage, Pawson et al (2004, p.15-16) conclude that, “although it is essential to clarify at some stage which of these approaches will drive the review, it may not be possible to make a final decision until the review is well underway. Certainly, we counsel strongly against the pre-publication of realist review ‘protocols’ in which both the review question and the purpose of the review must be set in stone before the real work begins.”

- **Articulate key theories to be explored**

For this stage Rycroft-Malone, et al. (2013, p.3) note that reviewers search for the relevant theories in the literature, outline, group and categorise or synthesize programme theories, design a theoretically based evaluative framework to be ‘populated’ with evidence and develop bespoke data extraction forms.

- **Search for and appraise evidence**

This stage involves deciding and defining the purposive sampling procedure to be adopted. It is at this stage that the reviewers define search sources based on factors such as accessibility, convenience and cost. The search terms and methods to be used (including cited reference searching) will be considered at this stage. It is vital that reviewers consider various terms that are used in reference to the phenomenon under study.

In our hypothetical example of CCE, one would find that scholars use a mix of terms to refer to climate change, these may include climate variability, global warming, climate fluctuation etc. The search terms do not appear from nowhere, they are suggested based on reviewers' background knowledge of the phenomenon to be studied (Alderson, Green, and Higgins, 2004). Reviewers will pick the most common terms to use in their search protocol. Finally, reviewers will decide on how they will reach searching saturation. This stage also includes the piloting of the search protocol to gauge its rigour and its relevance to the theory under test.

- **Extract and synthesise findings**

This stage involves the extraction of data to populate the evaluative framework with evidence. It also involves comparing and contrasting findings from different studies. The reviewers seek both confirmatory and contradictory findings. The programme theories are then refined in the light of the findings from the analysis of study data (Rycroft-Malone et al. 2013, p.3).

- **Develop narrative**

This final stage involves the coming together of the evaluators and other decision makers in the review of findings. Upon satisfaction, they will disseminate the review with findings, conclusions and recommendations. The objectives of the review will guide how the report will be disseminated (See also Module 5 on Communicating ESD Evaluation Findings). Academic findings will be published in academic journals, while practice-based reviews will be disseminated to practitioners for them to frame and reframe their programmes accordingly. Pawson et al (2004), however, noted that the aim of realist evaluation is not necessarily an instrumental one, that the review should lead to an immediate change in a given programme. Though that happens sometimes, a realist review is more inclined towards helping policy makers and practitioners 'make sense' of the way they understand and interpret the situations they encounter and the interventions they deploy.

What are the limitations and how these can be overcome?

In theory, realist evaluations can be carried out in many different circumstances, and alongside other types of evaluation. However, practically the method cannot be conducted by one person and it requires skills and knowledge of both the process and the type of intervention being evaluated (Higgins & Green, 2006).

Thus, a realist evaluation should not be conducted because the project has the attributes that are best evaluated by realist evaluation, without considering the human and material resources required for a meaningful evaluation.

Westhorp (2014) outlines various circumstances where realist evaluations are not the ideal approach to evaluation:

- A realist evaluation is not needed if an organisation already understands how, when and where a particular type of programme works. Realist evaluation is not particularly well suited for tried and trusted modes of delivery such as delivering nutrition programmes.
- A realist evaluation is only appropriate when stakeholders want to understand how and why an intervention worked, not when they are simply interested in knowing the project outcomes. In such a situation, results-based alternatives should be used.
- If a realist evaluation is to be credible and useful it requires good data on outcomes (or impact). This is because a realist evaluation seeks to generate a better understanding of what works and what doesn't work in different circumstances. This means it is essential that the intended outcomes (or impact) of the intervention being evaluated are properly specified. It also means it is important that these outcomes (or impacts) have either been assessed beforehand, or can be assessed during the evaluation.

Underlying theory

Realist evaluation is grounded in the philosophy of realism (Bhaskar, 2010). It also draws on systems theory. (For social programmes involving humans, as well as bio-physical realities, a particular kind of systems theory, called complexity theory, is often useful). Westhorp (2003, p.3) explains that the assumption is that everything in this world is organised in systems, which in turn are embedded in larger systems and connected to other levels. Social systems are open: elements can move in and out of the system. For example, family members can leave, others can marry into the family. Everything is embedded into other levels and all the systems interact with each other. As a result, any event has many causes, and at the same time may have many consequences. This also means that every outcome of a programme is a result of multiple causes; and that every programme may have many different outcomes. A realist

approach allows us to identify demi-regularities in the underlying generative mechanisms that contribute to observable outcomes in social-ecological spaces (see Bhaskar, 2010 and Pawson et al., 2004).

References for Realist meta-evaluation method

- Alderson P, Green S and JPT Higgins (Eds) (2004). Cochrane Reviewers' Handbook 4.2.2 [updated March 2004]. In: The Cochrane Library, Issue 1. John Wiley, Chichester.
- Bhaskar, R. (2010). Contexts of interdisciplinarity and climate change. In R. Bhaskar, C Frank, KG Hoyer, P. Naess & Higgins J. Parker (Eds). Interdisciplinarity and climate change: Transforming knowledge and practice for our global future. Routledge, London and New York.
- Green S. Eds. Cochrane Handbook for Systematic Reviews of Interventions 4.2.6 [updated September 2006]. In: The Cochrane Library, Issue 4, 2006. John Wiley, Chichester.
- Pawson, R., Greenhalgh, T., Harvey, G. and K. Walshe (2004). Realist synthesis: an introduction. ESRC Research Methods Programme, University of Manchester.
- Rycroft-Malone, J., McCormack, B., Hutchinson, A. M., DeCorby, K., Bucknall, T. K., Kent, B., Schultz, A., Snelgrove-Clarke, E., Stetler, C. B., Titler, M., Wallin, L., and V. Wilson (2012). Realist synthesis: Illustrating the method for implementation research. *Implementation Science*, 7(1), 33. <https://doi.org/10.1186/1748-5908-7-33>
- Stern, E. (2015). Impact Evaluation: A Guide for Commissioners and Managers. Prepared for the Big Lottery Fund, Bond, Comic Relief and the Department for International Development.
- Westhorp, G., Prins, E., Kusters, C., Hultink, M., Guijt, I., and J. Brouwers(2011). Realist Evaluation: An overview. 22.
- Westhorp, G. (2014). Realistic Impact Evaluation. An Introduction. A Methods Lab Publication, ODI. September, 2014.
- Wong, G., Greenhalgh, T., Westhorp, G., Buckingham, J., and R. Pawson (2013). RAMESES publication standards: Meta-narrative reviews. *Journal of Advanced Nursing*, 69(5), 987–1004. <https://doi.org/10.1111/jan.12092>

3B.5 Participatory Approaches

3B.5.1 Participatory evaluation

What is it about?

Participatory monitoring and evaluation (PM&E) involve the full participation of all stakeholders from the onset of the programme to the end (Guijt and Gaventa, 1998). These stakeholders could include, among others, intended beneficiaries, development agencies implementing the programme, policy makers and funders deciding together how progress should be measured and monitored, results communicated and acted upon, and then proceeding to do the evaluation together, as well. Participatory M&E offers new pathways of understanding and affecting inclusive change with the intended beneficiaries being at the center of the learning process. Besides its inclusivity, participatory M&E also aims to ensure accountability but not just to funders - to all stakeholders.

Guijt and Gaventa (1998) outline four broad principles at the heart of PM&E:

- **'Participation'** - this means opening up the design of the process to include those most directly affected, and agreeing to contribute to data gathering and do the analyse together;
- **Inclusiveness** - this requires 'negotiation' to reach agreement about what will be monitored or evaluated, how and when data will be collected and analysed, what the data actually means, and how findings will be shared, and action taken;
- **Learning** - is a key goal of PM&E and result of the above two processes; this shared learning then becomes the basis for subsequent improvement and corrective action;
- **Flexibility** - is essential, since the number, role, and skills of stakeholders, the external environment, and other factors change over time.

When is it used and what is its value?

Campilan (2000) noted that participatory evaluation is not a universally appropriate approach. It can only take place within the broader framework of a participatory development programme. It is incompatible with

linear, top-down research or development approaches because they operate under different assumptions and principles (see Module 4, for example, on criteria for validity). The exclusive use of external evaluators in a participatory development programme, for example, would be a gross contradiction.

For Campilan (2000), by involving all stakeholders (e.g., local people, collaborating organisations, programme field staff), participatory evaluation attains the following:

- A more well-rounded perspective of the programme being evaluated
- Support from a broader base of stakeholders and access to their knowledge, expertise and resources.
- Wider ownership and sharing of responsibility
- By allowing intended programme beneficiaries the space to represent their ideas and values, and have these recognized as significant, participatory evaluation promotes self-determination and agency amongst stakeholders.
- Validity of evaluation is enhanced through the multiple sources being tapped.
- Evaluation becomes ethically sound since it involves those who are most directly affected by its outcomes.

The epistemological argument for participatory processes of knowledge creation has been introduced in Module 3A, with reference to Santos (2018), who argued that we are living in a 'cognitive empire' in which only certain (modern Western scientific) knowledge forms are regarded as credible. According to Santos, this leaves everyone impoverished, because all knowledge is partial, and a much wider range of ways of knowing (including modern Western science) is needed to understand (for example) sustainability problems and how to respond to them. Participatory research and evaluation methodologies can help to expand the range of knowledge forms ESD practitioners and stakeholders have access to (establishing rich 'knowledge ecologies' that compensate for their partial natures), provided the evaluation methodology allows for the recognition of these 'other ways of knowing'. Of course, they are not guaranteed to do that, e.g., if all participants share the same understanding or perspective, the participatory process may not contribute to knowledge ecologies.

ESD example

The following hypothetical example of participatory evaluation is based on the guidelines provided by Aubel (2004) and Estrella & Gaventa (1998). The example involves farmers learning more about environmental degradation so as to improve their preparedness in addressing various existing and potential farming and livelihood challenges. How could such a learning process be evaluated in a participatory manner?

Step 1: Planning the PM&E process and determining the objectives and indicators

This stage involves the identification of stakeholders. Besides the farmers this could include their families and other community members, extension officers, development agencies, policy makers and funders. The objectives and each participant's role in the M&E process are discussed. Through a lengthy process of negotiations, contestations and compromises, indicators will be identified, along with the reasons for their inclusion and how and when they will be measured and by whom. The indicators may include food security, community knowledge, evidence of soil erosion, dam water levels, diversification of income streams, etc. Depending on stakeholders' interests, a common set of indicators are developed, however, where there are varying concerns and objectives, different sets of indicators should also be developed by different stakeholder groups (i.e. a mix of common and contextual indicators).

Stage 2: Data gathering

This stage involves the collection of all relevant data, which can be both qualitative and quantitative. Questions that should guide the conduct of data collection at this stage:

- What would be the source of data?
- Which tools will be used?
- Who should do data collection and when?



Quantitative data collections methods can include: community surveys; intercept interviews; and observations of the environment based on the chosen indicators. Qualitative methods can include various participatory learning methods using visual, interviewing and group tools and exercises.

Stage 3: Data analysis

Traditionally, data analysis is reserved for experts, however, under PM&E all stakeholders are involved at all levels. The farmers will therefore also be involved in drawing conclusions and in suggesting recommendations for best practices. Data analysis techniques to be used will depend on the data collection tools used and the type of data obtained in the previous stage (Estrella & Gaventa, 1998).

Stage 4: Documentation, sharing information and deciding on the way-forward

In as much as the preceding stages can strive to attain full participation of all stakeholders, this is not always feasible. This stage involves the sharing of insights and outcomes including with stakeholders who may not have taken part in preceding activities. Further discussions will be conducted on the actions to be taken to address emergent concerns and recommendations will be put forward. This stage requires great communication skills that include clarity, simplicity, accessibility, consideration of intended audience, precise messaging, and often also the use of non-verbal communication aides such as charts, images and maps (Estrella & Gaventa, 1998). Finally, stakeholders will pick the recommendations according to their priority and feasibility.

What are the limitations and how can these be overcome?

According to Campilan (2000), the following are common limitations that are associated with participatory evaluations and how they can be addressed:

- There is no consensus amongst evaluation practitioners on what a “true” participatory evaluation should look like.
- Participatory evaluation has the potential to disrupt existing power relations amongst stakeholders and if not addressed well, this can create more challenges than solutions. Participatory evaluation leads to contestations on various evaluation aspects such as objectives, criteria, measures and methods. Programme managers and supporters may disapprove of a participatory approach because of the perceived threat to their power and authority once they share with local people the control over the evaluation process and outcomes.

- Participatory M&E is often a challenging process for all concerned since it can require participants to examine their assumptions about what constitutes progress, and to face up to the contradictions and conflicts that can emerge.
- While a participatory approach seems ideal, in many instances intended beneficiaries may choose not to participate. This could be because they do not see the results as having direct and practical use for them. Deliberate efforts are needed to ensure that evaluation has concrete value for the stakeholders. Otherwise, participatory evaluation can be experienced as an unnecessary burden, and/or a token gesture by researchers in the name of participatory development.
- Evaluation is often experienced as an add-on responsibility by already busy programme staff. Hence the evaluation processes should be carefully designed to support their participation. For more on this, see developmental evaluations (Patton, 2011) and a summary of a participatory evaluation designed and implemented by AWARD (2018).
- Evaluation practitioners may struggle with participatory evaluations if it has not formed part of their training or prior experience. The method requires combined capacities in evaluation and in participatory approaches (see AWARD, 2018 and Rosenberg and Kotschy, 2020).
- Evaluation, participatory or otherwise, is a costly process in terms of money, effort and time. Yet these costs of evaluation are often not factored into programme planning and budgeting.

Underlying theory

Kemmis and McTaggart (2005) note that there have been four generations of participatory evaluation and the extensive history of the methodology cuts across many fields of social practice. The first generation began with the work of social psychologist Kurt Lewin (on action research).

Participatory research is driven by the values of social and environmental justice. It gives priority and a voice to the most disenfranchised members of the society. These may include the “victims” of injustices who may not have the capacity to bring scientific evidence of the problems that bedevil their communities as a result of insufficient influence and resources (Cargo & Mercer, 2008, p.329).

The postcolonial literature provides an important argument for participatory approaches to knowledge creation, as evident in the work of Santos (2018) quoted in this module, among many others, in recognition

of the epistemic injustices incurred when only certain kinds of knowledge were recognized as useful for addressing societal problems.

References for Participatory evaluation method

- Aubel, J. (2004). Participatory Monitoring & Evaluation for Hygiene Improvement. Beyond the toolbox: What else is required for effective PM&E? A Literature Review (Strategic Report 9). Office of Health, Infectious Diseases and Nutrition Bureau for Global Health, USAID. http://www.ehproject.org/PDF/Strategic_papers/SR-9%20Lit%20Rev.pdf
- AWARD (Association for Water and Rural Development) (2019). Harnessing monitoring and evaluation for learning. Experiences from the RESILIM-O Program 2019. www.award.org.za
- Cargo, M. and S.L. Mercer (2008). The Value and Challenges of Participatory Research: Strengthening its practice. *Annual Review of Public Health*, 29(1), pp. 325–350. <https://doi.org/10.1146/annurev.publhealth.29.091307.083824>
- Campilan, D. (2000). Participatory Evaluation of Participatory Research. Forum on Evaluation of International Cooperation Projects: Centering on Development of Human Resources in the Field of Agriculture. Nagoya, Japan, International Potato Center. <https://web.archive.org/web/20140703095000/http://ir.nul.nagoya-u.ac.jp/jspui/bitstream/2237/8890/1/39-56.pdf> (archived link)
- Guijt, I. and J. Gaventa (1998). Participatory Monitoring and Evaluation: Learning from change. IDS Policy Briefing. Brighton, University of Sussex. <http://www.ids.ac.uk/files/dmfile/PB12.pdf>
- Kemmis, S. and R. McTaggart (2005). Participatory action research: Communicative action and the public sphere. In N. Denzin & Y. Lincoln (Eds.) *The Sage Handbook of Qualitative Research* (3rd ed.). Thousand Oaks: Sage, New York.
- Rosenberg, E. and K. Kotschy (2020). Monitoring and evaluation in a changing world: A Southern African perspective on the skills needed for a new approach. *African Evaluation Journal*, Vol 8, No 1, DOI: <https://doi.org/10.4102/aej.v8i1.472>
- Santos, BdS. (2018). *The End of the Cognitive Empire. The coming of age of epistemologies of the South*. Duke University Press, Durham and London.

3B.5.2 Reflective practice as evaluation method

What is it about?

Reflective practice has been one of the most popular professional development and organisational learning theories that dissects the nexus between professional practice and continuing education programmes, and can also be recognized for evaluation purposes. For Kinsella (2010), the theory is concerned with understanding professional knowledge. It seeks to answer the question:

- What does it mean to know in professional life?
- How might we conceive of professional knowledge for the good of the practitioner, the learner (if the practitioner is an ESD facilitator), the professions and the field (of, in this case, ESD), and society more broadly?

Tay and Jain (2019) explain that practitioners like educators encounter unexpected challenges and novel phenomena in their daily work, and without well calculated interventions, artistry and skills, their work may be impossible. Schön (1993) puts forward an explanation on how practitioners consciously seek to understand these challenges and the reasons a routine intervention may suddenly not be working anymore. The process he described is termed 'reflection-in-action', through which practitioners progressively devise new ways of addressing unanticipated challenges (Cameron, 2009).

When is it used and what is its value?

Over the years, reflective practice has been integrated into professional induction programmes, continuing education programmes, and by professional bodies of varying professions that include health and social care (Kinsella, 2010). Depending on the profession and ideological background that informs the reflective practice to be used, McLaughlin (1999, p.11), notes that there are two questions that should be asked before one jumps on the "bandwagon":

- ✓ How 'reflection' is to be understood. These include queries about the nature of reflection (e.g., how explicit and systematic it is or should be, and how it is related to action) and about its scope and objects (e.g., the matters on which teachers are invited to reflect about).

- ✓ Second, questions about the value of reflection, include whether 'reflection' is an outcome on its own or a means to an outcome. In education, McLaughlin (1999, p.11) asserts that, it is better put in the following form: is reflection valued simply as a process, or in terms of the quality of the judgements- and possibly action - to which it leads?

ESD Example

This section will provide a hypothetical ESD example of reflective practice:

Farmers take part in a training workshop on the use of environmentally friendly ways of farming. The knowledge and skills that these farmers get exposed to will be practically relevant to the desired environmental outcomes. The knowledge and skills may include avoiding streambank cultivation, pest control, use of hazardous chemicals and cultivars that may be detrimental to environmental ecosystems, etc. Upon the completion of the training workshops farmers are awarded with certificates and are regarded as knowledgeable on matters to do with environmental sustainability and it is now their duty to put into practice (in varying contexts) the information that they have acquired. However, after a couple of seasons with plausible outcomes, a disaster strikes. A rampant locust plague that has the potential to wipe away all the crops and a drought with the potential to dry out all their crops. Farmers are now faced with an unprecedented dilemma which they try to understand and solve through reflective practice. There is a discrepancy between what they learnt (theory) and what they are experiencing (practice). Tay and Jain (2019), labeled these two distant fronts as "sloppiness" (theoretical and common problems) and "hard zones" (unpredictability of practice). Behind these two fronts, is the clash between professional values, goals, purposes of their knowledge and skills and the new reality that was at least not part of their training and at most condemned in their training.

In the case of the locusts, farmers may use chemicals that will control the locust but will have long-term environmental effects. In the case of drought, farmers may choose to sink more boreholes to improve their water supply, however this may threaten their local aquifers. To maneuver around this dilemma, farmers go through an evaluative introspective process of asking themselves question:

- What does it mean to be a farmer today and in the future
- What would be intended and unintended consequences of their preferred interventions.

This continuous process will require time, objectivity and the understanding of one's self in relation to their positionality in their new farming realities and in other future possible realities.

The limitations and how these can be overcome

Though reflective practice has gained much recognition and has been widely adopted, there remains are concerns about the lack of conceptual clarity surrounding the term reflective practice and the notion of reflection itself. Kinsella (2010). The notion of reflection has become paradoxical and now has subjective meanings such that it is sometimes used on concepts that have nothing to do with reflexivity and the term reflective practitioner is in some cases used loosely without proper bases for its use. McLaughlin (1999) explains that the term has become fashionable because of its political and educational convenience, especially in opposing emerging trends and reforms in the education sector.

To put this into perspective, McLaughlin asks:

- Who, after all, would want to champion the unreflective practitioner?

This question exposes the fluidity of reflective practice. One reason this is so, is because Schön's theory of reflective practice is open to such a wide range of interpretations (Kinsella, 2010).

To address these dilemmas McLaughlin (1999) proposes two solutions:

- First, there is a need for a shift from sloganeering about the reflective practitioner and giving a proper definition to the concept. This is achievable through exploring the concept in detail to find out what it means and what it involves as well as the extent of its influence in both training and training of the trainers.
- Second, there is a need for a shift from the concept in its traditional form towards acknowledging its shortcomings. Thereafter, it should be strengthened and be strategically positioned in a well-defined and reliable context that would ensure a universal description of what it means.

Underlying theory

Reflective practice is a continual learning strategy that is rooted in the philosophy of constructivism. Central to both constructivism and reflective practice is that knowledge and beliefs are formed subjectively within learners and cannot be transmitted without the active involvement of the learner. The continued development of knowledge (learning) builds on prior knowledge and grows out of experience, and specifically problematic experience, reflective practice expands on constructivism, with its distinction between two types or levels of cognitive activity: the theory-in-use and the espoused theory (Osterman, 1998). In the same way that reflective practice recognizes prior learning or preconceived ideas, constructivism recognizes how an individual incorporates the new experience/insights into the existing knowledge without discarding it. Thus, newly acquired knowledge simultaneously exists in one's mental framework (Adom, Yeboah & Kusi Ankrah, 2016).

References for Reflective practice as evaluation method

- Adom, D., Yeboah, A., and K. Ankrah (2016). Constructivism philosophical paradigm: Implication for research, teaching and learning. <http://www.eajournals.org/wp-content/uplo>
- Cameron, M. (2009). Review Essays: Donald A. Schön, *The Reflective Practitioner: How Professionals Think in Action*. New York: Basic Books, 1983. ISBN 0—465—06874—X (hbk); ISBN 0—465—06878—2 (pbk). *Qualitative Social Work: Research and Practice*, 8(1), 124–129. <https://doi.org/10.1177/14733250090080010802>
- Kinsella, E. A. (2010). Professional knowledge and the epistemology of reflective practice: The Epistemology of Reflective Practice. *Nursing Philosophy*, 11(1), 3–14. <https://doi.org/10.1111/j.1466-769X.2009.00428>
- McLaughlin, T. H. (1999). Beyond the Reflective Teacher. *Educational Philosophy and Theory*, 31(1), 9–25. <https://doi.org/10.1111/j.1469-5812.1999.tb00371.x>
- Osterman, K. F. (1998). Using Constructivism and Reflective Practice to Bridge the Theory/Practice Gap. *Bridging the Gap*, 19.
- Tay, P. L., and J. Jain (2019). Learning to Teach Through Reflective Practice. *International Journal of Education, Psychology and Counseling*, 4(30), 342-355.

3B.6 Complexity-sensitive Approaches

3B.6.1 Developmental Evaluation

What is it about?

Developmental evaluation is built into the DNA of an innovative programme; its role is to track what emerges under conditions of complexity, documenting and interpreting the dynamics, interactions, and interdependencies that occur as the programme unfolds, and working with the implementers to make sense of what emerges. In this method for evaluating educational or development programmes, the evaluation process also adapts as the programme adapts to contextual changes and new insights.

At different stages of the programme cycle, the evaluation plays different roles. It helps to identify what the programme goals and objectives (and possibly the theory of change) are, and keeps track of how it may change over time, as in complexity-sensitive programmes and evaluations, there is an understanding that greater specificity about goals and objectives will develop over time, and that the theory of change may not only become clearer during implementation, but may even change due to contextual changes, or changes in understanding based on feedbacks during implementation, or both.

Evaluation is central in providing the feedbacks through ongoing monitoring. The evaluation team's understanding of what is to be monitored, also changes over time.

Developmental evaluation is suitable for programmes whose implementers are social innovators and change agents and who, (with the support of their funders and other stakeholders) keen to collaborate with the evaluator(s) to conceptualize, design, and test new approaches in a long-term, on-going process of development, adaptation, and intentional change. The evaluator or evaluation team has an integral role in the programme, which includes elucidating team discussions with evaluative data and logic, and to facilitate data-based decision making in the developmental process as the programme unfolds (Smith, 1994b, p.220)

How does it work?

Developmental evaluation is in some ways a general approach to evaluation, and various other methods could be described as developmental as well. However, as a specific method for evaluation, the evaluator joins a programme team very early on in the process, to also help with programme design, with a built-in evaluation dimension. One or more members of the evaluation team is usually stationed within the programme and work alongside programme implementers to monitor agreed-upon aspects of the programme, gather additional information from time to time, share findings, and convene programme implementers to reflect on findings and consider what they mean.

As the programme adapts to changing circumstances and/or insights, so the evaluation itself also needs to be open to adaptation. Over the life cycle of the programme, the evaluation also needs to adjust to the programme's changing nature and needs. At the start, for example, when the implementers are very busy getting implementation activities off the ground, they may be less amenable to spending long hours in reflection meetings.

When is it used and what is its value?

Developmental evaluation is useful in innovative settings where goals are emergent and changing rather than predetermined and fixed, time frames are fluid (e.g., in longer term programmes) and the purposes are innovation, change, and learning rather than (or in addition to) external accountability (Patton, 2011, p.viii).

Accountability is first and foremost to the implementers themselves, as opposed to external parties like funders, as it is assumed that the implementers driving the programme, as social innovators and change agents, have very high stakes in their programme achieving what they broadly set out to do.

Developmental evaluation is suitable for complex contexts because it is sensitive to the context, and it helps to track contextual changes and reflect on their implications for the programme's theory of change and implementation plans. Responsiveness and adaptation, both to the context and to the life cycle of the programme, are key features of developmental evaluation. It is especially appropriate for situations of high

uncertainty where what may and does emerge is relatively unpredictable and uncontrollable. A recent example of such a situation is the Covid-19 pandemic; others may be innovations such as new approaches to community development, or experiments with green economy initiatives.

Programmes using adaptive cycles or strategic adaptive management, find developmental evaluation most aligned to their way of work. The evaluators need to be comfortable with “uncertain beginnings, muddled middles and unpredictable endings that ripple on and on without end” (Patton, 2011, p.9) and to understand these as unavoidable features of innovative change-oriented projects in complex social systems.

ESD example

In a partnership between USAID and AWARD (the Association for Water and Rural Development) in South Africa, the evaluators were brought in to help with programme and evaluation design early on. The five-year programme (eventually extended to 7 years) was seen as an innovation to build resilience to climate change and other stressors in the coupled social-ecological systems in a large landscape, the Olifants River catchment. The project involved multiple water and land user groups, from isolated rural villagers and subsistence farmers to game farms, conservation areas, commercial farmers, mining companies and the government agencies responsible for providing services and protecting natural resources in the face of devastating droughts, deep social and economic inequality, good governance risks and other challenges. As AWARD rolled out its programme of diverse capacity building and support initiatives, starting with a deep contextual analysis and stakeholder engagement, the evaluation team worked alongside them to decide on suitable indicators, map out broad and specific theories of change, and adapt reporting formats to suit the stage in the life of the programme, as well as its changing feedback and learning needs. Implementers were closely supported by the evaluators, who helped to convene reflection meetings, analyse field reports, and co-author quarterly and annual reports. Extensive data was needed for ongoing strategic adaptive decision-making, and ranged from case studies using most significant change stories (in the early stages of the programme) to value-creation evaluation methods and indicator-based quantitative monitoring in the middle and later stages of the project. (For more on this project see <http://award.org.za/>).

Underlying theory

Developmental evaluation as described by Patton (2011) is informed by systems theory and complexity theory in general, and in particular by the theory of complex adaptive systems (CAS). Principles for operating in complex adaptive systems therefore inform the practice of developmental evaluation. Characteristics of complex adaptive systems, to which this kind of evaluation aims to respond, summarized by Patton (ibid) are:

- nonlinearity
- emergence (see Module 3A)
- dynamical
- adaptive
- uncertainty (see Value creation method), and
- coevolutionary

While CAS theory is well known in the fields of ecology, natural resource management and sustainability sciences, in organisational contexts it has been extensively developed by Stacey (2001), which influenced the developmental evaluation design promoted by Patton (2011) described here. Patton's description of developmental evaluation also draws on the learning theories developed by Argyris and Schön (1978), on single and double loop learning and reflective practice (see Reflective Practice in this Module). In single loop learning, implementers compare the difference between what they aimed to achieve and what they achieved, and adjust their actions to better reach their target. In double loop learning, implementers go further and interrogate the system that led to the problem in the first place, and aim to make system changes so that the cause of the problem can be addressed, and to embed the solution in a changed system (Patton, 2011, p.11).

Limitations and how they can be overcome

Developmental evaluations work best where there is an individual or preferably, a group of people in the organisation or setting being evaluated, who care about the evaluation and the findings it generates. Under these circumstances, there is optimal participation and findings are used. However, programme

implementers often regard evaluation as an intrusion into their busy implementation schedules, and or a judgement on their performance. It takes time to and patience to build trusting relationships with all members of the implementation team, numerous orientation sessions, and a sensitivity to and willingness to adapt to the implementers' needs, to build a strong team consisting of both evaluators and implementers, all keen and interested in what the evaluation processes can offer the implementation processes.

Another limitation is of course that stakeholders want to see evaluation results in recognizable formats, often quite early on in the life of a long-term programme, and developmental evaluations can use any of the other methods described in the module to share a mix of quantitative and qualitative findings as they become available, with the caveat that they represent feedbacks rather than final findings.

(See in particular Process-oriented approaches: Value-creation method).

References for Developmental evaluation method

- Argyris, C. and D. Schön (1987). Organisational learning: A theory of action perspective. Addison-Wesley, Reading.
- Patton, M.Q. (2011). Developmental Evaluation: Applying complexity concepts to enhance innovation and use. The Guilford Press, New York.
- Stacey, R.D. (2001). Complex responsive processes in organisations: Learning and knowledge creation. Routledge, New York.

3C CONCLUSION AND CONNECTION WITH OTHER MODULES OF ESD EVALUATION CURRICULUM

There are many more formal evaluation methods that could have been included here. There is also more nuance and detail about each method that should be known by anyone wanting to use them to practically design and undertake an evaluation. The Module is neither comprehensive in its coverage of all formally recognised evaluation methods, nor offering a detailed handbook on how to conduct evaluations using the methods covered. It is aimed midway between introducing a useful spectrum of methods and associated approaches, and giving enough information about them to help course participants decide which to explore further. References are provided for such further exploration. The Module provides some important practical pointers, but practical details should be worked out further in a particular context, with reference to any of the resources that are specific to the chosen method, or generic but suitable for the chosen method and context. The unique contribution of this Module is in situating the available methods in the context of ESD, and commenting on their suitability for this context, as informed by the orientation to both ESD outlined in Module 1, and the ‘game changing’ orientation to evaluation and ESD described in Modules 1 and 2 and the accompanying reading (O’Donoghue et al., 2019).

Throughout Module 3 there has been reference to the underlying assumptions that shape particular evaluation methods, or the use of these methods. These are sometimes described as different evaluation or research paradigms, but the field of evaluation is in practice fluid, featuring various links, lineages and overlapping trends, so that it may be difficult or impossible to capture it in discrete paradigms (on this point, see also Pawson and Tilley, 2007). Course participants engage again with this idea in Module 4.

Course participants may conclude that the chosen methods reflect “qualitative” rather than “quantitative” evaluation and this is to some extent true, if by qualitative we mean “not empiricist”. However qualitative and quantitative are in some ways labels most suitable to refer to different types of data. Most methods introduced in this Module can make use of both qualitative and quantitative data, but with the exception of the experimental design, the different data forms are not used with an empiricist assumption that quantification is the only way in which to really make sense of the findings. For example, in the Most Significant Change, participants may share quantitative information in their “significant accounts”, e.g. the growing size of the area of land which they replanted with trees, could be a significant change story for

them. However, the evaluation conveners would not necessarily be looking for the most frequently mentioned change to decide which is the most significant.

Rather than calling them “qualitative methods”, the notable features of the majority of evaluation methods and approaches in Module 3 are:

- the turn towards inclusivity and participation (see also Module 2) and
- valuing diverse knowledge forms;
- designing for complexity, and
- designing for real time learning as part and purpose of evaluation processes.

Module 3 gives participants the language to reflect on the different approaches they may want to explore for their own evaluation designs. In the process of design, one encounters the need to consider how best to ensure that the evaluation findings will be regarded as credible or trustworthy by the stakeholders, ranging from intended beneficiaries, implementers and funders to broader publics, as well as other ESD practitioners eager to know how to improve their own practices and programmes. To that end, evaluation design also includes a significant consideration of validity criteria, associated with the different approaches to and methodologies for evaluation, and this is discussed in Module 4. Module 5 engages another important consideration, already raised in this Module, which is how the findings of evaluations are to be communicated. Module 5 will show that communication strategies, too, may differ depending on the underlying approach, method, as well as purpose(s) for a particular evaluation.

3D ASSIGNMENTS



ASSIGNMENT 3.1: Coming to Grips with Different Evaluation Methods & Approaches

Compare and contrast two evaluation methods in terms of the following:

1. How they are done (typical sequence or processes)
2. Typical scope and purpose served
3. Their underlying assumptions
4. Contexts in which they are commonly used
5. Historical development
6. Credibility considerations (validity)
7. Typical tools/instruments and activities

Presentation Format

- ✓ Participants to produce two roleplays filmed and followed with a text file or recorded presentation.
- ✓ Uploaded to e-learning platform
- ✓ Feedback according to given rubric

ASSIGNMENT 3.2 Describe purpose of ESD evaluation

Describe an ESD context of your choice, and the purpose of evaluation required in that context. Then, using the course content, decide with reasoned justification which method(s) could be used in such a context and for such a purpose. Using the template provided, design an ESD related evaluation, to more deeply explore the use of a particular evaluation approach and method. In certain formats of the course, following Module 4 on communicating ESD evaluation, students will implement their design (with support); reflect on the outcomes; and use the experience to re-design the evaluation process to apply, deepen and consolidate their learning.

- ✓ Students produce a written outline with implementation detail of the proposed evaluation (proposal format)
- ✓ Uploaded to e-learning platform (pre-implementation stage) for feedback

GUIDELINES TO DEVELOP ESD EVALUATION PROPOSALS

Describe an ESD context of your choice, and the purpose of evaluation required in that context. Then, using the course content, decide with reasoned justification which method(s) could be used in such a context and for such a purpose. Using the template provided, design an ESD related evaluation, to more deeply explore the use of a particular evaluation approach and method. In certain formats of the course, following Module 4, students will implement their design (with support); reflect on the outcomes; and use the experience to re-design the evaluation process to apply, deepen and consolidate their learning.

1. **Title of Project**
2. **ESD Context**
3. **Stakeholders in the evaluation of this ESD Context**
4. **Purpose of the evaluation**
5. **Main evaluation question/s**
6. **Evaluation Method to Use**
7. **Justification for this Method**
8. **Processes, instruments and/or tools to use to gather information for the evaluation**
9. **Analytic tools and / or framework to use**
10. **Key resources and references that will be consulted**

QUALITY and VALIDITY for ESD EVALUATION

With Matthias Barth, Eureka Rosenberg, Rob O'Donoghue

the problem of
COMPLEX
LANGUAGE...

what is the
BIG PICTURE?

how can we PROPERLY
evaluate ESD?

how can it be
more INTEGRATIVE?

are UNIVERSITIES
ready for this kind
of EVALUATION?

we need
EVALUATORS with
an OPEN HEART and
the WILL to change
-and EXPERIENCE-

ESD EVALUATION
SUMMERSCHOOL
19.08.21

VALIDITY

can we TRUST
the EVALUATION?

how USEFUL
is it?

(post-empiricist)

does it CATALYZE
TRANSFORMATION?

would others
CONFIRM it?

and to
WHOM

building TRUST
from the very
BEGINNING

who are the
PEERS?

**MEMBER
CHECKING!**

how can we
really work TOGETHER?

PEER REVIEW

involvement of
different STAKEHOLDERS

what to do when
nobody participates?

how to
measure it?

can we have a
MIXED MASALA?
using different (empirical) methods

ALL levels of
perception matter

it's a
PROCESS!

how do we define
LEARNING?

time & timing!

it's about
AUTHENTICITY!
& being clear!

what do you want to MEASURE?

what questions
do you need?

= QUALITY

and what
instruments?

PURPOSE — QUESTION — DESIGN

pick the
"LOW HANGING
FRUITS"
they are rich
in DATA

looking in
the PAST
to understand
the PRESENT
and shaping
the FUTURE

Photo credits @ISDL Leuphana



QUALITY AND VALIDITY FOR ESD EVALUATION

JUAN CARLOS A. SANDOVAL RIVERA
EURETA ROSENBERG
MATTHIAS BARTH
ROB O'DONOGHUE

3E Quality and Validity in ESD Evaluation (Sub section of Module 3)

By Juan Carlos A. Sandoval Rivera

LEARNING OBJECTIVES

After successfully finishing this sub section on quality and validity of Module 3, the participants are able to:

- Justify the choice of evaluation approach, informed by a panoramic knowledge of the main validation and reliability approaches that can be applied in evaluation processes, based on the main epistemological paradigms existing in the social sciences.
- Explain the main characteristics of the different paradigms and their credibility considerations.
- Support their decisions regarding the validation criteria that will support their evaluations, to have an acceptable reliability regarding the findings that are generated during and at the end of the evaluation process according to the purpose and context in the one that is done.
- Find relevant ways to answer, bringing answers to questions and reflecting on trustworthiness and authenticity of the answers.
- Participants will be able to find relevant ways to answer, bringing answers to questions and reflecting on relevance of the answers

The purpose of the sub section on is to show the different paradigms on quality and validity that currently exist in the social sciences and how in each of them are different criteria to validate the evaluation processes in which the course participants are involved. Participants will be able to identify the paradigm that best suits their purposes and will be able to position themselves in the paradigm of their choice to generate criteria that can validate the evaluation questions and findings they make.

The two main considerations are:

- ✓ Justification of the choice of evaluation approach, according to the selected paradigm

- ✓ Learning to support decisions regarding the validation criteria that will support the evaluation processes in order to have an acceptable reliability

Terminology

- Paradigms
- Quality
- Validation

Epistemological paradigms in Social Sciences

Currently, in a general way, three paradigms coexist in the social sciences: the positivist, the interpretive and the critical paradigm. These paradigms, arising from consolidated theoretical perspectives, have dissimilar epistemological and, therefore, methodological assumptions, hence the progress and reflection produced within one of them cannot be applied easily to the rest. Likewise, these paradigms are frequently at the base of the approximations used by social scientists to account for social reality. The coexistence of paradigms is not an exception, but the rule in the social sciences and nowadays, no longer generates significant controversies. The acceptance of such coexistence is linked to the need to use different methods, embedded in these various paradigms, more to capture the complex and multiple nature of reality than to guarantee the validity of the results obtained or, in other words, to deepen more in the analysis and make it more complex than to seek objectivity.

Below we will briefly explain the most relevant characteristics of the three paradigms that are coexisting in the social sciences.

Positivist paradigm:

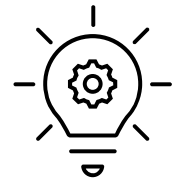
- External observation of social phenomena: The researcher is not involved in the process he/she observes.
- Causal explanations: They seek to explain what are the causes that produce certain situations.
- The verifiability and / or refutability of theories: It is aimed at analyzing whether the data obtained prove that a theoretical approach is correct, or if, on the contrary, the data refutes it.

Interpretive paradigm:

- The need to understand the meaning of social action in the context of the life world and from the perspective of the participants.
- The understanding of the symbolically pre-structured reality of each context requires the participatory work of the researcher, who does not "give" meaning to what is observed but makes explicit the meaning "given" by the participants.

Critical paradigm:

- Considers that knowledge is always built by interests that start from the needs of human groups; It aims to liberate the oppressions to which human beings are subjected and is achieved through participation for social transformation.
- Its main assumptions are to know and understand reality from practice and action; unite theory and practice, integrating knowledge, action and values; direct knowledge towards the emancipation and liberation of the human being; and propose the integration of all participants, including the researcher.



ASSIGNMENT 3.3 – Different approaches in Quality and Validity

Compare and contrast the different paradigms, analyze what their main characteristics are as well as identify (in connection with formal methods for ESD Evaluation) what are the different approaches and methodologies related to each paradigm, scope and purpose served and tools/instruments and activities that are used within each paradigm.

Validation and Quality in Social Research

"Quality" within the social sciences, regardless of the type of research, was evaluated mainly based on four criteria:

- ✓ **the internal validity of the data**, implied verifying whether they correctly reflected the unique and independent external reality of the various perspectives that could evaluate it;
- ✓ **external validity**, if the statistical inference of the measurable characteristics of the sample allowed us to know the population parameters;

- ✓ **reliability**, if it guaranteed the stability of the findings regardless of the researcher and the moment;
- ✓ **objectivity**, if the knowledge referred to the object, and not to the biases and / or prejudices of the researcher.

"Validity" redefined as credibility implies "reflecting on the credibility or correctness of the knowledge produced and adopting strategies to achieve it" (Maxwell, 2012)¹. It means being able to assess confidence, both in the result of the study and in its process. The researcher is the fundamental instrument of the investigation, and although this immersion in the field is what guarantees the validity of his data, it can also constitute a danger if a critical distance is not taken from the process under examination.

As a first approach, the following procedures are suggested to achieve validity in a study:

1. adopt a commitment to field work,
2. obtain theoretically rich data,
3. data triangulation,
4. review by the interviewees, and
5. review by peer and external researchers.

Validation Criteria

Below are some procedures that can help validate a research project and thereby sustain its quality, which were listed in the previous section:

1. **Commitment to field work:** perform it responsibly, observing and recording information for as long as necessary; record the various views of the participants that will support the conclusions; write exact, complete and precise notes, differentiate between the original data and own interpretations.
2. **Obtaining rich data:** detailed, dense and complete information that can lead to a greater understanding of the studied situation.

3. **Triangulation:** is a strategy to increase “reliability” in the quality of the data to be analyzed; This strategy is used to recognize the limitations that a single data source, view, or method implies to understand a social issue.
4. **Review by the interviewees:** also called communicative validation, it implies stating that the interviewees will be asked to critically read the various research documents, so that they can evaluate the quality of the descriptions, the survey of all perspectives and the capture of their meaning.
5. **External reviewers to the research team:** relevant procedure within the validity process; the study, before being published, is evaluated by peers with academic background and practitioners.



ASSIGNMENT 3.4 VALIDATION AND QUALITY IN SOCIAL RESEARCH

Participants develop their own validation criteria based on the 2 readings (Lincoln, Lynham & Guba, 2018; Flick, 2007) that are relevant to their evaluation processes and present them through an orally commented PPT file. This file will be uploaded to the platform so that it can be compared with the criteria developed by the other participants and participate in a group discussion through a virtual forum.

References for sub section for Quality and Validity in ESD Evaluation

- Flick U “Managing Quality in Qualitative Research”M
- Denzin K. N & Lincoln S.Y “The SAGE Handbook of Qualitative Research”
- Video: <https://www.youtube.com/watch?v=f0oQ3a9ATbM>

DIALOGUE ON QUALITY AND VALIDITY FOR ESD EVALUATION



Quality and Validity for ESD Evaluation: Video

In this [video](#) recorded at the Summer School on “Regenerative ESD and Evaluation for the Future We Want” the speakers Eureka Rosenberg, Matthias Barth and Rob O’Donoghue discuss the role of quality and validity for ESD Evaluation

Please click on the link above to access the video

Below is the summary of the dialogue between Prof. Matthias Barth, Prof. Eureka Rosenberg and Prof. Rob O’Donoghue, on quality and validity for ESD Evaluation during the summer school on "Regenerative Education for Sustainable Development and Evaluation for the Future We Want" on dated 19 August 2021 in a virtual format. The full transcript of the dialogue is can be seen in later parts.

Five Key points

1. Knowing whether a university or a funder or any other important evaluation partner are ready or not for evaluations is an important step as sometimes organization aren’t prepared for a particular kind of evaluation. Others suggest that you plan for an evaluation by doing a feasibility study to see the readiness of an organization and what needs to be done to prepare it to be ready to participate freely and openly in an evaluation in such a way that everybody will learn from the process and make the ESD initiatives better in the process because that’s what one wants to do.
2. The quality and validity of an evaluation begin around how we start, not when things are over when we ask, “was this a quality study and wasn’t valid?”. This means the involvement of different stakeholder voices, not only in the end as the receiver of the academic wisdom that was brought in an evaluation, but as partners who co-develop knowledge throughout an evaluation process. However, this requires the evaluation to be set up in such a way that people are going to want to participate in it and to build trust between the different partners in the evaluation so that even when some challenging information come to the fore, people would feel comfortable to work with it and try and do something with it as opposed to shutting it down. The worst is an evaluation that never

sees the light of day because the partners perceive the evaluation as doing much harm, even if it's done in a proper way. That's why evaluations can take quite a long time, and the developmental evaluation is quite helpful because one can slowly prepare the partners in the organization to be ready for it.

3. Validity refers to the extent to which the findings are plausible, credible, and trustworthy. That is, you still need to be able to defend the outcomes of your evaluation, even if you are not basing your evaluation on measurements. It would be best to justify that using qualitative data and an open-ended evaluation, a participatory process, one that is not based on indicators, one that doesn't rely heavily on measurements, one that considers things that can't be measured and still value them as necessary. If you can communicate that, then the organization and the partners should consider that as still a proper evaluation.
4. Member checking with the participants is a common way of making sure that what you as the evaluator come up with is shared and valued and trusted, and found credible by the other participants in the evaluation. Member checking in ESD evaluation is more often in the form of a conversation with members, rather than just checking for validity. However, this should not be done in an imperious manner. Member checking is connected to confirmability - to what extent do others confirm that, yes, indeed, we also conclude that this program is worth continuing. That can do that by laying a careful data trail and also doing peer review.
5. Questionnaires need a lot of attention in terms of the validity of their design. This also involves internal validity, which asks how careful you are in implementing your evaluation. External validity which refers to whether the findings can be generalized to contexts outside the one in which you've done the evaluation. So, you can do it very, very carefully, but because of the way you set up your design, you can't generalize. Content validity will refer to the scope of that questionnaire - are you asking a spectrum of questions that reflects the scope of what one might call ESD competence, or are you only narrowly asking knowledge questions? To what extent do the questions you're going to ask and the answers you're going to get from that question refer to the construct of ESD competence?



Formal Evaluation Methods (Module 3)

Presented by Prof Eureka Rosenberg

Chair: Environment and Sustainability Education
South Africa



PROGRAMME

1. Orientation to Module 3
Time for Questions

- 2. Examples unpacked:**
- Experiments, Surveys
 - Narrative methods
 - Process focussed methods
 - Participatory methods
 - Theory-based methods
 - Complexity sensitive methods
 - Mixed methods / multi-modality and critical realism

3. Discussion

4. Assignment
Time for Questions



Assignment 3.2 – Design an ESD Evaluation (Proposal)

Describe an ESD context of your choice, and the purpose of evaluation required in that context. Then, using the course content, decide with reasoned justification which method(s) could be used in such a context and for such a purpose. Using the template provided, design an ESD related evaluation, to more deeply explore the use of a particular evaluation approach and method. In certain formats of the course, following Module 4, students will implement their design (with support); reflect on the outcomes; and use the experience to re-design the evaluation process to apply, deepen and consolidate their learning.

1. Title of Project
2. ESD Context
3. Stakeholders in the evaluation of this ESD Context
4. Purpose of the evaluation
5. Evaluation Method to Use
6. Justification for this Method
7. Main evaluation question
8. Evaluation sub-questions
9. Processes, instruments and/or tools to use to gather information for the evaluation
10. Analytic tools and / or framework to use
11. Key resources and references that will be consulted

A Range of Evaluation Approaches

Constructivism, interpretivism; narrative: Example Most Significant Change Story method (Davis & Dart)

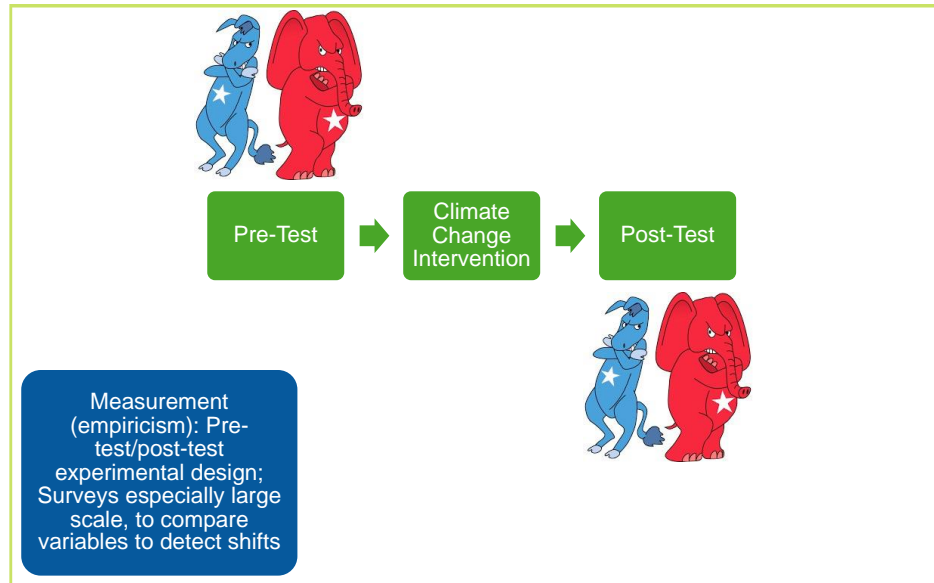
Process-based: Appreciative enquiry; Developmental evaluation (MQ Paton), Value creation framework (Wenger & Trayner)

Participatory: Participatory evaluation, action research (Lewin); Reflective practice (Argyris & Schon)

Measurement (empiricism): Pre-test/post-test experimental design; Surveys especially large scale, to compare variables to detect shifts

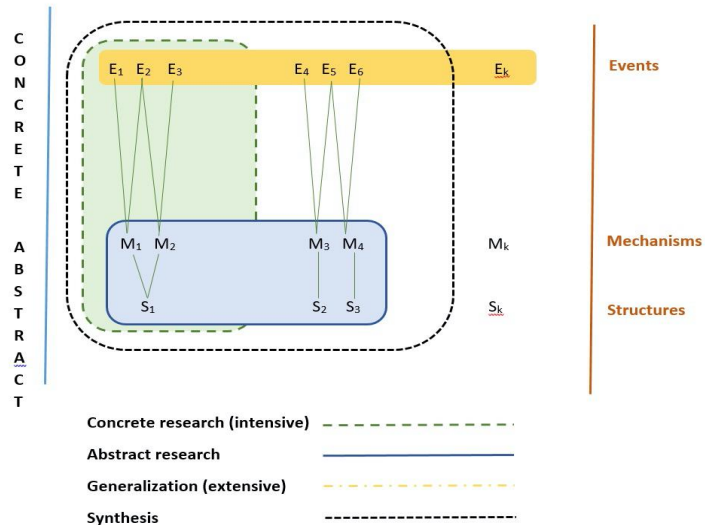
Theory-based : Designing a theory of change; Realist evaluation; Realist meta-evaluation/ synthesis

Measurement-based Approaches



ESD Evaluation Course

(From Sayer, A., 2010, *Problems of Explanation and the Aims of Social Science*)

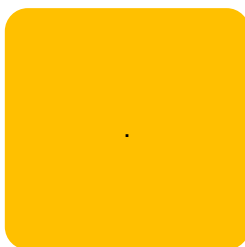
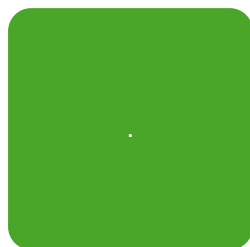


Narrative Approaches

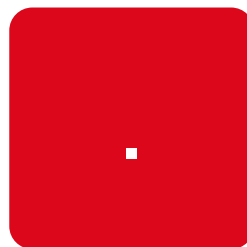
Constructivism,
interpretivism;
narrative:
Example Most
Significant
Change Story
method (Davis &
Dart)

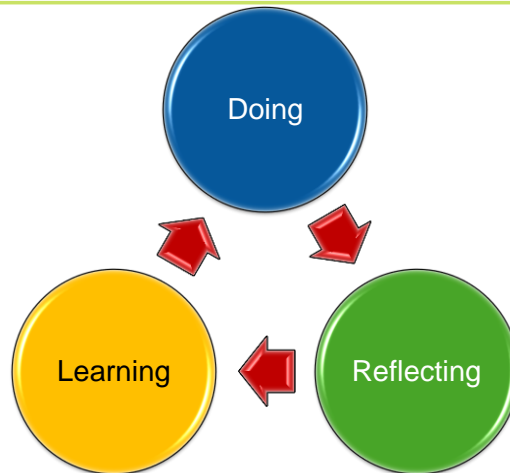


Participatory Approaches



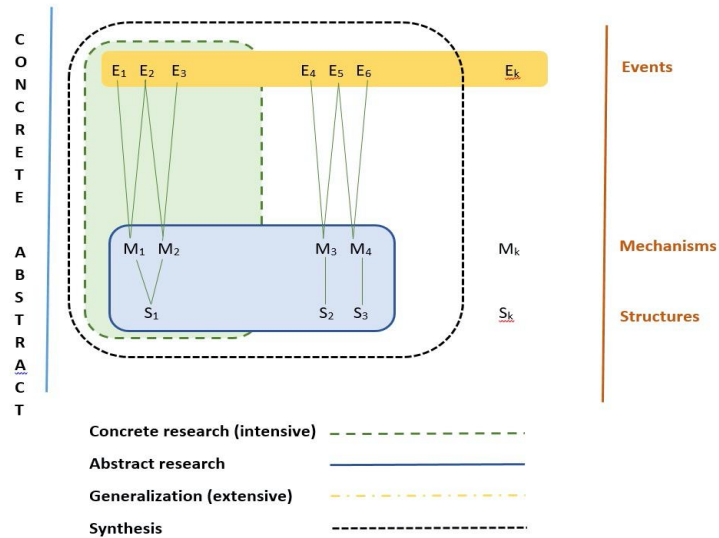
Participatory:
Participatory
evaluation, action
research (Lewin);
Reflective practice
(Argyris & Schon)



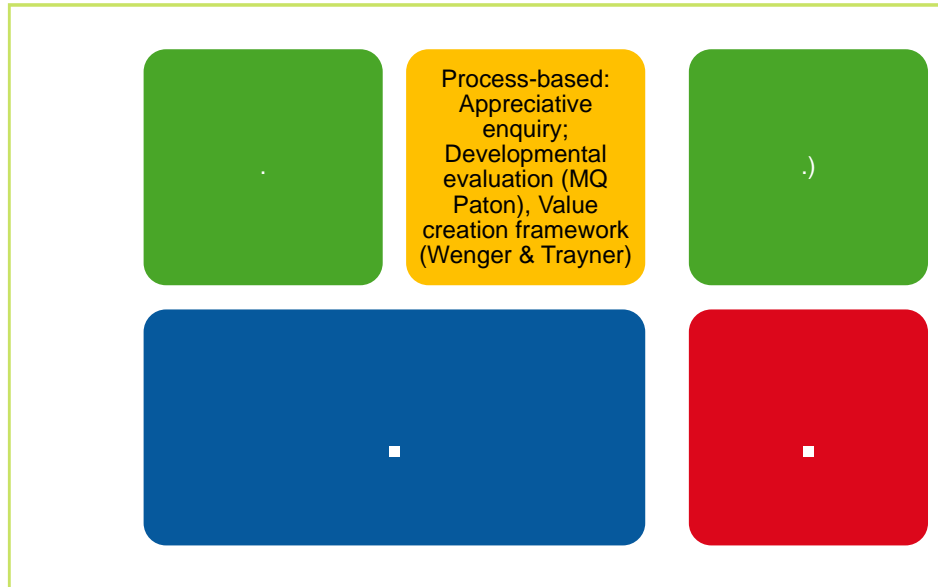


Reflective practitioner approach (Argyris & Schön)

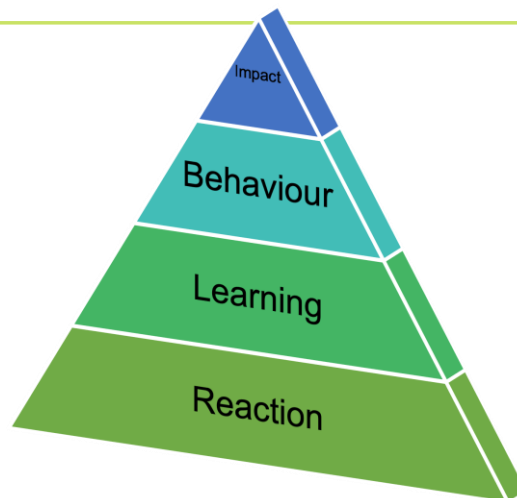
(From Sayer,
A., 2010,
*Problems of
Explanation
and the Aims
of Social
Science*)



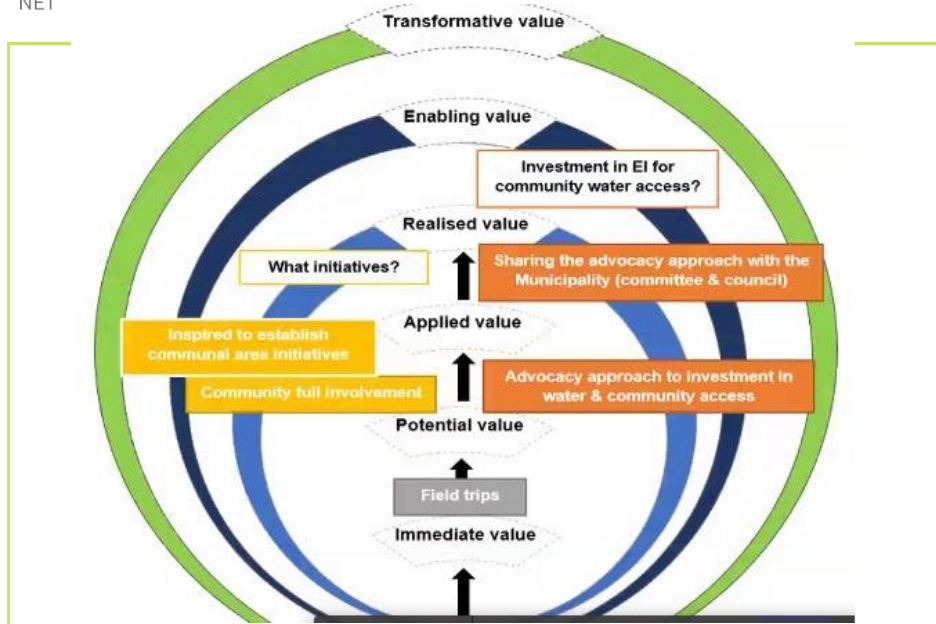
A Range of Evaluation Approaches



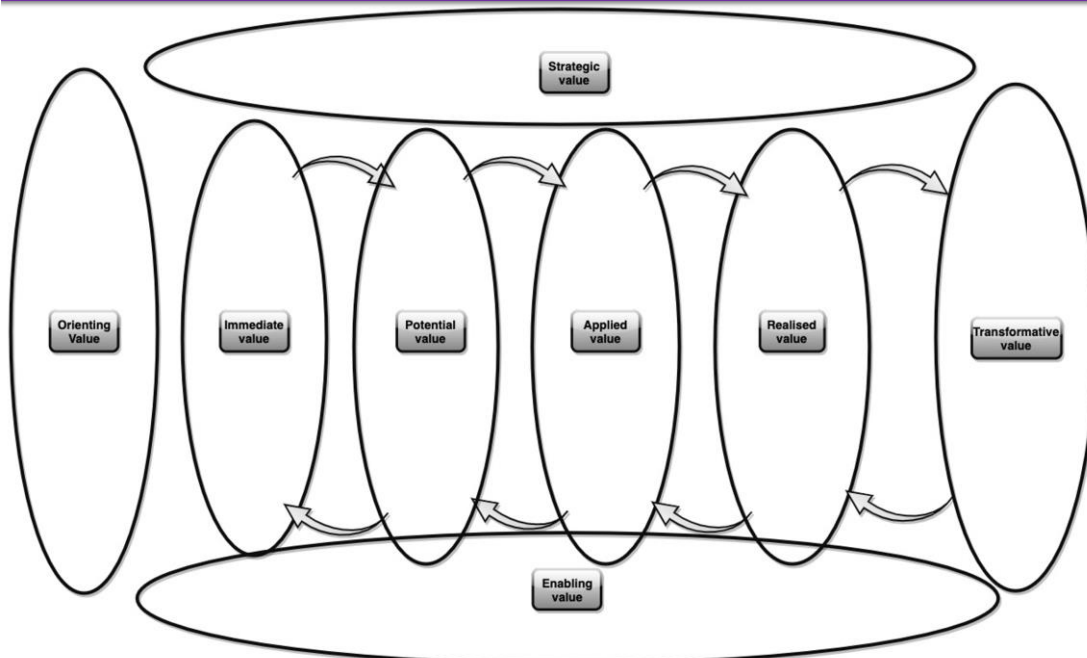
ESD Evaluation Course



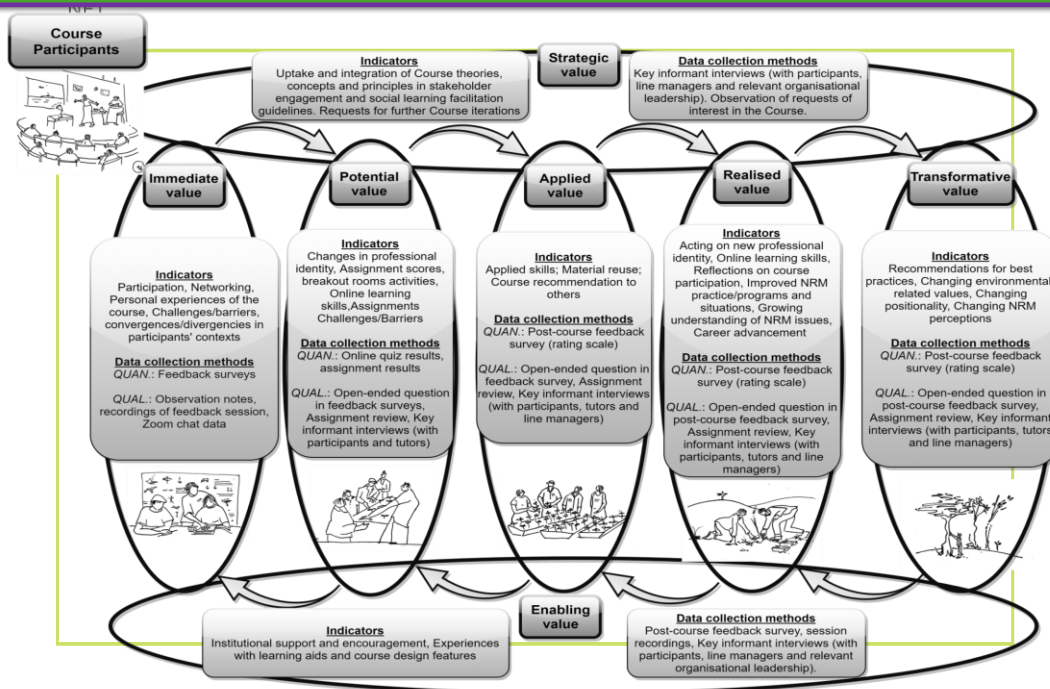
Kirkpatrick, D.L. 1994. *Evaluating Training Programs: The four levels*. San Francisco, Berrett-Koehler.



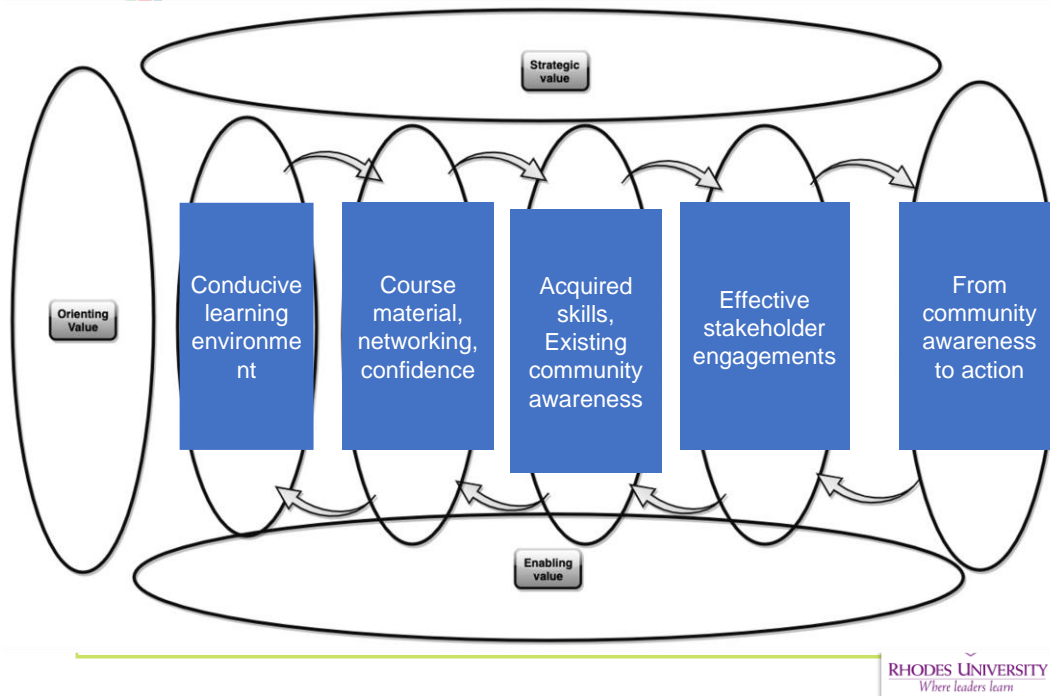
A Framework for Capturing Value Created during an ESD Programme



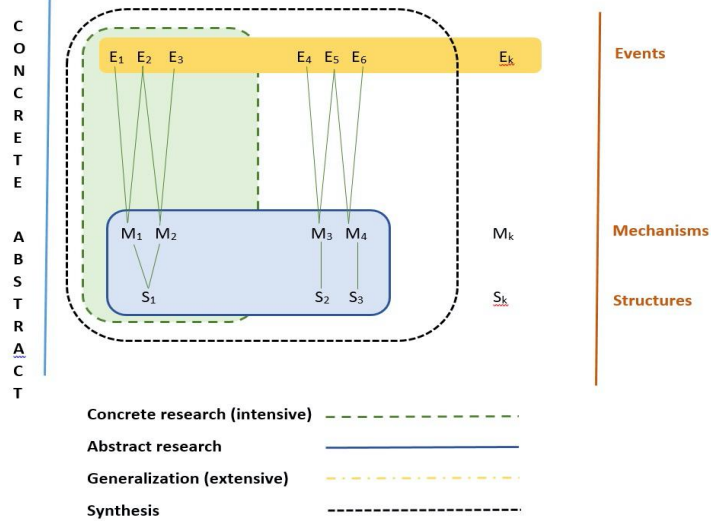
Value Creation Cycles (Processes of participation)



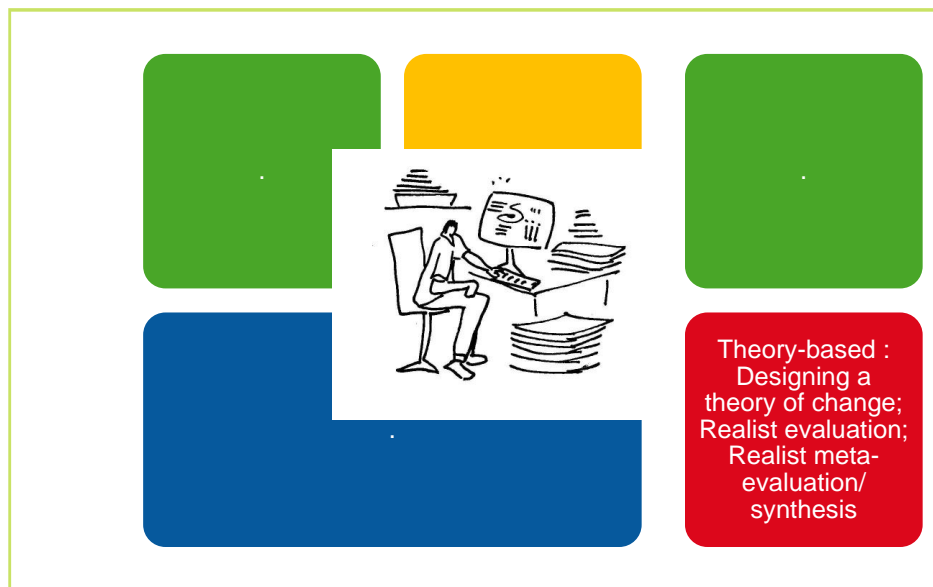
Example: Value Created for Students



(From Sayer,
A., 2010,
*Problems of
Explanation
and the Aims
of Social
Science*)



Theory-based Approaches



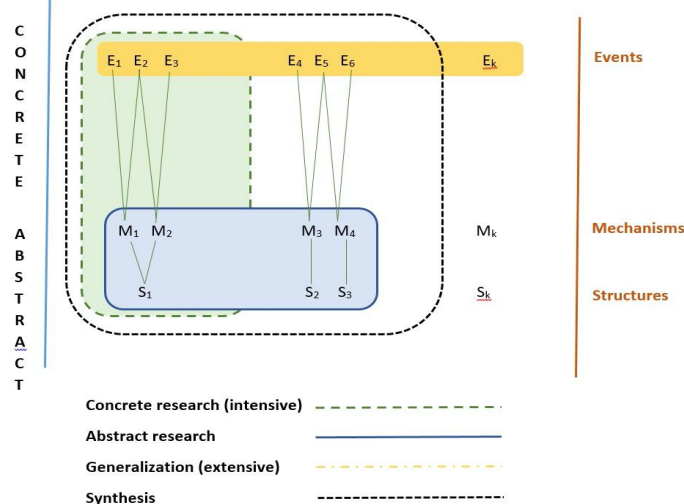
Assignment 3.2 – Design an ESD Evaluation (Proposal)

Describe an ESD context of your choice, and the purpose of evaluation required in that context. Then, using the course content, decide with reasoned justification which method(s) could be used in such a context and for such a purpose. Using the template provided, design an ESD related evaluation, to more deeply explore the use of a particular evaluation approach and method. In certain formats of the course, following Module 4, students will implement their design (with support); reflect on the outcomes; and use the experience to re-design the evaluation process to apply, deepen and consolidate their learning.

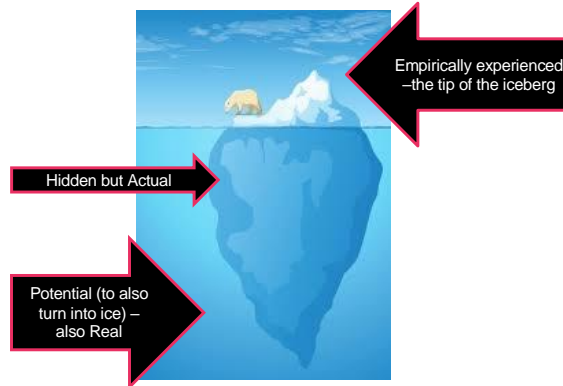
1. Title of Project
2. ESD Context
3. Stakeholders in the evaluation of this ESD Context
4. Purpose of the evaluation
5. Evaluation Method to Use
6. Justification for this Method
7. Main evaluation question
8. Evaluation sub-questions
9. Method to use
10. Processes, instruments and/or tools to use to gather information for the evaluation
11. Analytic tools and / or framework to use
12. Key resources and references that will be consulted

Critical Realism and Multi-Modality

(From Sayer, A., 2010, *Problems of Explanation and the Aims of Social Science*)



Accessing layers of reality



Critical & social realists (Bhaskar, Archer, Sayer) argued that reality is stratified
A depth ontology allows evaluators to investigate more than what can be empirically observed, to also look at what reality must be like, for the empirical to appear. This includes looking for *absences* - what should be there but is not, and why? (*critical realism*)



RHODES UNIVERSITY
Where leaders learn

Over to You – Questions? Comments?

Transcription of Dialogue on Quality and Validity for ESD Evaluation

Speakers: Prof. Eureka Rosenberg, Prof. Matthias Barth and Prof. Rob O'Donoghue

Date: 19 August 2021

Audience: Participants of Summer School on ESD Evaluation

Platform: Zoom

Duration of Video: 35 minutes

Link of Video: Please click [here](#)

Matthias: Yeah, welcome everybody. Good to see you again. We have a nice session for after lunch, I think that is perfect. I'm always looking forward to these sort of discussions because they sound so theoretical and indeed they are so practical when it comes to, you know, the very question of, you know, how do we ensure quality in evaluation research? I think that is tremendously important and we want to frame that a bit first, if you give a bit of an input and then really apply that to your work in smaller groups and bring that together. I think this resonates well with what you've done so far, but Rob is much better equipped than I am to justify that because he's been following much more closely than on what happened the last two days.

Rob: That's what we thought we'd start up with... was a little bit of a recap around some of the questions that arose when Matthias and I were listening to the start around the ESD as an evaluative process, the question of transformation and the question of regeneration that occupied our minds quite considerably. So, Marcus and Juliane started by saying are our university is ready for evaluation, as we have been exploring it through the online course. And then, the question that always comes up is that it's associated with complex language and scientific approaches and it's often very difficult when people come to ESD evaluation from different levels and from different disciplines and approaches. And then Mirka really opened it up really nicely when she talked about project evaluation and indicators. And she talked about the concept of delivery and the different demands that would relate to quality and reliability. So, the question that we raised is, how can it be evaluated properly? How can we properly evaluate ESD? What would be proper evaluation?

And then came up the concept. Well, you know, what is the endgame here? What is the big picture? How do we get around the paradigms and the mental blocks that are associated with this kind of plural, disciplinary work? And this, the different levels that are associated with it and then came the question of the professionalisation of evaluation and this leads to an externalization that Eureta already has addressed with us.

What does it mean to be part of evaluation, to have this more inclusive approach to evaluation in ESD and how do we navigate across professional evaluation approaches, which are more externalized and being part of an evaluation as an inclusive process? The navigation of epistemological and ontological foundations that are producing different perspectives. And how do we build common understandings that Thomas raised opening up yesterday evening for us, the idea -morning for some or afternoon for others- the idea that evaluation is latently normative. What does the normativity of evaluation mean to us? Because normativity kind of implies a bringing together that needs to be both reliable and predictable and also promote agreement and some sort of progress around particular understanding of quality and reliability. So, we thought that would sort of set the scene out of the course for this conversation that we're both about to have and then we thought well, if one of the questions is our university is ready for evaluation, and we thought we'd start with Eureta saying, you know, from a university perspective of evaluation research. What are the ranges of discourse on quality and reliability that we could usefully be engaged with? So over to you, Eureta.

Eureta: Yeah, so it's a very busy slide, but it's going to come in some small bits. So I won't, I won't take long with this and it's only this one slide, but it's speaking to the idea of validity in evaluations and I'll make mention of reliability as well. And the idea of the quality, the quality of the evaluation. What are the quality criteria by which our evaluations are going to be judged, and by which we are going to judge them ourselves. So, whether University or a funder or any other important evaluation partner, whether they are ready or not for these evaluations is a very good question and sometimes an organization isn't ready for a particular kind of evaluation. And some people even suggest that when it starts and you're planning for an evaluation by doing a feasibility study to see how ready is the organization and what needs to be done to prepare it to be ready to participate freely and openly in an evaluation in such a way that everybody will learn from the process and make the ESD initiatives better in the process, because that's what one really wants to do. So, you can do a sort of a scan to determine the extent to which the organization is ready. Some of you will be floating your proposals past the others in the, who you might want to work with, and you can hear the responses but you might already be anticipating that they won't be ready for a participatory evaluation or a

more open-ended kind of evaluation that are not based very narrowly on objectives and indicators. Then how do you communicate to the other partners that, nonetheless, even though the form of evaluation that you want to use might be participatory, open-ended, indicator three, still it is a proper evaluation. So that's where the idea of validity comes in because validity is what we use to decide whether the knowledge that was created in this evaluation can be trusted and that it is meaningful, it can be used to further inform ESD processes. The research methodologist Patti Lather described the conversation about validity as between a rock and a soft place. Because on the one hand, you have the rock of the empiricist approach which is entirely based on quantitative data and statistical analyses. And in this case, the validity is the extent to which the measurements are accurate, and the conclusions are accurate and the findings are reliable. If you repeat the same evaluation, if another person comes in and does the same evaluation, they would find exactly the same findings. So that's validity in the empiricist framework.

Now to move on to the post-empiricist, what some would call qualitative framing for your evaluation, but rather you can use quantitative and qualitative data in a post empiricist framing. Here validity refers to the extent to which the findings are plausible, credible, trustworthy. You still need to be able to defend the outcomes of your evaluation, even if you are not basing your evaluation on measurements. And if you can do that, if you can justify that using qualitative data and an open-ended evaluation, a participatory process, one that is not based on indicators, one that doesn't rely heavily on measurements, one that actually takes into account things that can't be measured and still value them as important. If you can communicate that, then the organization and the partners should be able to consider that, as still a proper evaluation. So, one thing to say is that I know that many of you want to become professional evaluators and I'm very glad to hear that because the field really needs well-informed evaluators with open minds and willingness to contribute to change and evaluation can contribute to change but only if we think carefully about the methodologies that we use.

You can't become a professional evaluator from doing this course. It's not comprehensive enough. You need to combine the course with a lot of practical experience and a lot of further reading. What this slide is doing and what our entire course is doing, in our discussion is going to do today is just to give you some orientation to the reading. So, when you encounter all of these terms in the literature face validity, content and construct validity, internal and external validity and a whole host of post-empiricist forms of validity, the important thing is that you need to, you will find all of the definitions for all of these in the literature. Don't

worry. I don't need to give you the definitions, you can find them. What I give you and Matthias and Rob give you is the framing with which to make sense of what you'll be finding in the literature. So, look at construct and content validity, for example, these are forms of validity that apply when you're doing an instrumentalist study or you are making use of very strong measurement-based study. For example, if you're going to use a questionnaire to determine ESD competencies. Content validity will refer to the scope of that questionnaire. Are you asking a spectrum of questions that really reflects the scope of what one might call ESD competence or are you only narrowly asking questions about knowledge? Maybe you're not including anything related to agency or values. Then I could say the content validity of your question is not great because you are only asking a narrow set of questions.

Construct validity, on the other hand, would refer to the most important construct, in that question would be the construct of ESD competence. It's a theoretical idea that you've constructed and to what extent do the questions that you're going to ask and the answers that you're going to get from that question actually refer to the construct of ESD competence. These are some of the questions you must ask yourself when you design questionnaires. Questionnaires really need a lot of attention in terms of the validity of their design. And then there's the internal validity, which is how carefully are you implementing your evaluation and the external validity refers to whether the findings can actually be generalized to contexts outside the one in which you've done the evaluation. So, you can do it very, very carefully, but nonetheless because of the way you set up your design, you can't actually generalize. For example, if you've got a small scale very context (13:30) case study. Then there's no external validity. So those are considerations when you are busy with an empiricist type of evaluation. On the other hand, if you look at the post-empiricist studies, here, one of the validity criteria, the principles for the quality evaluation would be how useful is the report or whatever it is that you're producing at the end of the evaluation. That's a very important Criterion. You can have a perfect laboratory setting, beautifully executed, but it might not be particularly useful because you can't really apply the findings to any other situation. Then the utility is low and it might not be all that valid. So the credibility, the trustworthiness, how do you improve that? Member checking with the participants is a common way of making sure that what you as the evaluator come up with, is actually shared and valued and trusted and found to be credible by the other participants in the evaluation. Confirmability: to what extent do others confirm that, yes, indeed, we also conclude that this program is worth continuing, confirmability. And you can do that by laying a careful data trail and also doing peer review. I'm nearly finished. I know this is taking a bit of time.

I want to share with you just this idea of catalytic validity, because we've spoken so much about instrumental ESD versus emancipatory transformative approaches to ESD education and sustainability.

The same Patti Lather who talks about between a rock and a soft place, it comes from a critical standpoint in her research and talks about the catalytic validity as the degree to which the research process in our case, the evaluation process, and reorient, and focus, and energize, the participants in what Paulo Freire called conscientization. To what extent does this evaluation help us to know reality in order to better transform it? Because that is the intention transformative. So, to be able to do that, one needs to recognize, not only the reality, altering impact, potential impact of the research or evaluation process itself, but also the need to consciously channel this impact so that respondents gain self-understanding, ideally self-determination or agency through research participation. So here you would be looking for something quite different from whether the measurements were accurate, but rather the extent to which the participation in the evaluation process, have helped people to better transform the program and other aspects of reality that are important. So this shows you the contrast between different approaches to validity and that's what you need to keep in mind when you go into the literature and you read more about this.

Yeah, let me pause there.

Rob: Thanks Eureka, that's really useful because I think it sort of sits up on both sides of the tennis court, you know, where you pass the ball between the two. But when you're playing at the net and you are at the cutting edge of things passing really fast and you have to be in the game, the question that kind of comes to mind is, can you have a mixed Masala? Okay, mixed methods as you showed us yesterday, where people are able to include empirical work with the partners that they are working with and probably has quite big implications for where does quality and validity begin. Now, to my mind, it's really important for us to think quality and validity begins when we start, and it begins around how we start, not when things are over and we say, was this a quality study and wasn't valid. That's what often happens. And I think that was really very usefully brought to the fall by Mirka, when she said, you know, properly. And I've had evaluations that have been judged, you know, was this done properly and who speaks to, you know, what is proper, and how do we actually deal with that?

Eureta: Can I respond, Rob? And it was also a question from Elaine, I think, is a useful one to look at, validity also applies to other research instruments, but particularly closed instruments such as questionnaires or tests, or assessments very much the internal construct and content validity would apply to them. Just to say that the difference between empirical and empiricist is very important here. Empiricist would imply that the only empirical data that I would value would be measurements, but I can use empirical data including measurements in a different more fluid open-ended way in a post-empiricist study. So, just to clarify that. That's why I'm not calling a quantitative versus qualitative, because that isn't really the distinction. I can use very freely work with the facts, figures and statistics, do surveys, gather a lot of information, big data, combine it with in-depth case studies, but I do so freely with what the critical realist school of depth ontology. In other words, recognizing that the layers of reality that are important to be able to understand ESD as a phenomena in a particular university, or school, or other context, the layers of reality that are important above the empirically observable, as well as people's feelings, perceptions, values, the underlying forces that gave rise to those underlying constraints that make things possible or not possible in a particular university, or company, and business and so forth. This picture here of the different layers of reality. Now, from a depth ontology point of view, in critical realism we say that all of those layers are important and you can bring them together in an evaluation to give you that full picture.

Rob: I like the way you raised the question of member checking, you know, so, who are the most important members for assuring quality and reliability? Now, do you get a person like Matthias and yourself who's got a lot of experience in academic research to look at what you're doing and help design it, so it's going to pass sort of master around the first question that we had? Is our university going to be able to accept this? Should the university be the mediator? To my mind, in a lot of work we can get the support of the university and do a good evaluation, but we're working with partners and we're working as people confronting climate change and wondering how we can do things together? How do we evaluate things at a local level? So it is quite a mixed challenge to get the balance right for our projects.

Eureta: Yeah. Well, I'm sure Matthias can share experience from Leuphana here, but I would think that if I want my evaluation to make a difference at my university, I need to set it up as you say, Rob, right from the start, set it up in such a way that people are going to want to participate in it and to build a trust between the different partners in the evaluation so that even when some challenging information might come to the fore people would feel comfortable to actually work with it and try and do something with it as opposed to

shutting it down. The worst is an evaluation that never sees the light of day because the partners are going -no, we don't agree with it and even if it's true, it's doing us so much harm, we don't want to hear that-. That's why evaluations can take quite a long time and the developmental evaluation is quite useful because one can slowly prepare the partners in the organization to be ready for it. Now, many of the participants on this course won't have that much time, but some will, some will have up to three years, so it's a very nice period of time to be able to work - a quick in-and-out type of evaluation, can sell them, you know, be particularly catalytic or help to build up an organization. There's also a nice question from Senan, and I'm still not knowing correctly about whether they are instruments to show catalytic validity, but we can come back to that later. Maybe, I think, Matthias wants to respond to this one. Oh wait...

Matthias: I just want to add to a bit and give you a slightly different framing because I think what you guys see here and this is really important is, you know, it's first the attempt really to try to translate sort of the traditional empiricist criteria to a post-empiricist approach. This is one. You will often find it in the literature, then there is, what does criteria one, two, three be in an post-empiricist way or literature, or quantitative versus qualitative? There you will find the differentiation. So, this is one approach we will come across. The second one, this is where catalytic validity comes in nicely, is, you know, what else do we need? You know, what is the additional, what brings an added value if we do an evaluation? So, it's not just a translation to bring in sort of into a new context but also to add to that. And one thing here with the catalytic aspect, you really have the transformative element, which plays an important role, especially when we think about emancipatory, educational approaches.

I want to bring in another additional nuance, which I find extremely helpful, and this is what lately came out of the formative evaluation in transdisciplinary research because that also adds a slightly different layer, really trying to come up with quality criteria, that do justice to the specific setting in a transdisciplinary evaluation. So, the involvement of different stakeholders of voices, not only in the end, you know, as the receiver of the academic wisdom that was brought in an evaluation but really as partners who do co-develop knowledge throughout an evaluation. So, you will find also literature on here and emphasis, especially on formative evaluation in transdisciplinary research, which I find really quite helpful because it gives you sort of a guideline also of what to think about in such evaluation approaches.

Rob: Yes, often when we think about quality and reliability and validity, you know, the academic perspective is peer review. My paper post-peer review is going to be published. You know, its past is sort of like a quality scrutiny, but in the work that we're doing on evaluation. Who are the peers? And what groups of peers do we need to take into account right from the start? You know, and how do we work with the peers that we have with colleagues and partners to co-develop a grasp of the situation and where the work is going? I always find it very useful to revert to a peer review and it was interesting the way Eureka pointed out in Patti Lather's work, how important the idea of member checking is, you know. Sometimes it appears in empiricist work, checked with the members, and they said that that's what they said. And now, you know, based on this analysis. This is what I think. So, I wonder often, if you can take member checking in ESD evaluation much more into a conversation with members, rather than just checking for validity. How do we treat and work with partners as peers? So, we are not always imperious, you, of being on the outside and being in the academic world, you know, it's the academy that mediates evaluation. I really like the way you discussed the point about the usefulness of evaluation, useful to whom? Useful to my career because now I'm getting a publication and I'm becoming an evaluation professional or is it really contributing to ESD in the situation in which I'm working in partnership with people, struggling in evaluative journeys to future sustainability? So, there's always a lot of to and from and a lot of thinking and conversations to have in this kind of work.

Matthias: Yeah, that was what I was going to say. So, there's the chance to jump into the conversation now, but really also to bring that conversation forward. Our suggestion would then be to split the group into smaller groups of three and really to then raise the question. Okay, If we look into these sort of frameworks and guiding criterias, what would that mean, if we take one of your examples? Now, going through the assignments in module 1, I've already seen so many rich examples, which we can perfectly use and I would like to see you as a group. Also take one of these cases and discuss. What would that mean in terms of quality assurance in an evaluation process? How can we do that? What consequences would that have?

About the Authors

Eureta Rosenberg is the Chair of Environment and Sustainability Education and Director of the Environmental Learning Research Centre at Rhodes University. Her qualifications include an M.Sc. in Medical Science (Hematology), an Honours in Psychology, and a Masters in Education with a focus on Environmental Education. Her PhD, which she obtained in 1995, focused on transformative research methodology. Her other scholarly interests include evaluation that supports organisational learning; strategic planning; research ethics suitable for a diversity of research traditions; skills for sustainability, and green work and learning. Eureta supervises M.Ed. and PhD students and Post-Doctoral Fellows in these and related research areas. She is a member of the Environmental Education Association of Southern Africa's Council and the Southern African Journal of Environmental Education editor-in-chief. At her academic home, Rhodes University, she serves as the Dean of the Education Faculty, Chair of the Education Faculty Research Ethics Committee, Chair of the Rhodes University Ethical Standards Committee, and member of the Institutional Planning Committee.

Ludwig Chanyau is a PhD student in the Environmental Learning Research Centre at Rhodes University and a researcher in the programme on Monitoring and Evaluation of Climate Change Education (MECCE) in partnership with the Sustainability and Education Policy Network. Ludwig holds a Masters in Social Development from the University of Cape Town, where he researched migrants' capabilities and access to social services. He later taught and supervised postgraduate Social Development studies at the same university between 2019 and 2020. Ludwig has also worked as a researcher in the development sector in Zimbabwe for UNICEF and in South Africa for local organizations working in youth development and urban food systems. His current research is on the value of climate change-related social learning among women farmers. He uses the Value Creation Framework (Wenger-Trayner and Wenger-Trayner, 2020) to evaluate green skills courses and support evaluation course within the ESD Expert Net.

Juan Carlos A. Sandoval-Rivera is a full-time researcher at the Instituto de Investigaciones en Educación, Universidad Veracruzana, Mexico. Through audio-visual collaborative research methodologies and qualitative research, Juan Carlos researches and supervises educational research projects on the interaction between indigenous knowledge and scientific knowledge in socio-ecological processes and adaptation to climate change in indigenous and rural contexts.

This module has been designed by team of student assistants Carolin Ellerkamp, Charline Rieffel, Edali Beltran, Johanna Philipps, Maxi Baumert with support from Deepika Joon, Ludwig Chanyau and Jennifer Krah.

The members of ESD Evaluation Working Group of ESD Expert Network contributed with their ideas towards the development of the module.

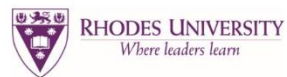
For any further query on the content of the module, please write to esd_evaluation@leuphana.de

If you are interested to join the Special Interest Group on Value Creation Framework at Environmental Learning Research Centre, Rhodes University, please write to esd_evaluation@leuphana.de

Supported by ENGAGEMENT GLOBAL



with funding from the Federal
Ministry for Economic Cooperation and
Development





ESD EVALUATION