LITERATURE REVIEW:

SUSTAINABLE OUTCOMES AND THE SYSTEMS FIELD

JANUARY 20, 2015

This publication was produced for review by the United States Agency for International Development. It was prepared by Management Systems International for the E3 Analytics and Evaluation Project.

LITERATURE REVIEW:

SUSTAINABLE OUTCOMES AND THE SYSTEMS FIELD

Management Systems International
Corporate Offices
200 12th Street, South
Arlington, VA 22202 USA
Tel: + 1 703 979 7100

Contracted under AID-OAA-M-13-00017

E3 Analytics and Evaluation Project

Prepared by:
Richard Hummelbrunner

DISCLAIMER
The author's views expressed in this publication do not necessarily reflect the views of the United States Agency for International Development or the United States Government.

CONTENTS

Acronyms and Abbreviations **2**

Executive Summary **3**
- Introduction 3
- Methodology 3
- Findings 3
  - Dynamic Sustainability 3
  - Understanding Sustainable Outcomes 4
  - Evaluating Sustainable Outcomes 4

Introduction **5**
- Audience 5

Methodology **5**

Findings **6**
- Part 1: Use and Understanding of the Terms "Sustainability" and "Sustainable Outcomes" 6
  - Definition(s) of Sustainability 6
  - Sustainability and Adaptation/Adaptability 9
  - Contributions from the "Systems" Field to the Sustainability Discourse 12
- Part 2: Evaluating Sustainable Outcomes 19

Conclusions **25**
- Implications for Ex-Post Evaluation Series 26

Annex A: Bibliography **27**

Annex B: Websites **31**

Annex C: Document Search Terms **32**

Annex D: Statement of Work **33**

ACRONYMS AND ABBREVIATIONS

CAS	Complex Adaptive Systems
CLD	Causal Loop Diagrams
DAC	Development Assistance Committee of the OECD
DANIDA	Danish International Development Agency
dTS	Development and Training Services
E3	USAID Bureau for Economic Growth, Education and Environment
LER	USAID/PPL Office of Learning, Evaluation and Research
MSI	Management Systems International
OECD	Organization for Economic Co-operation and Development
PPL	USAID Bureau for Policy, Planning and Learning
RUN	Three-step model defining the adaptability of social groups
SEC	Sustainability Evaluation Checklist
USAID	United States Agency for International Development
VSM	Viable Systems Model

EXECUTIVE SUMMARY

Introduction

USAID's Office of Learning, Evaluation and Research, in the Bureau for Policy, Planning and Learning (PPL/LER), is undertaking a series of ex-post evaluations (i.e., evaluations that take place several years after projects are completed) on 'sustainable outcomes' of USAID programs. The purpose of this evaluation series is to better understand what factors promote sustainable outcomes in development programs and projects.

This literature review is intended to inform the ex-post evaluation series by examining the evidence or evidence gaps related to factors that contribute to the sustainability of development program outcomes, in particular with respect to systems and complexity theory and thinking.

Methodology

Part 1 of this review is concerned with the use and operationalization of the terms "sustainability" and "sustainable outcomes" and the relationship between sustainability and adaptability. Part 2 identifies lessons learned from previous evaluations and proposes some designs and methods that are appropriate for evaluating sustainable outcomes.

The literature review draws on a range of documents, including academic articles, working papers, discussion papers, book chapters, case studies and relevant grey literature. The review was not intended to be a comprehensive search of all materials relating to the topics, but is based upon a curated selection of documents intended to be most helpful for designing this evaluation within the boundaries of sustainability and systems.

Findings

The review has revealed a diverse picture with respect to the understanding of sustainability and its evaluation. This is not only in the field of basic education, but also in social development at large. Beginning with the rather static definition by OECD/DAC, and some suggested improvements, there are different (and sometimes incompatible) ways of understanding the term "sustainability." However, there is also a recurrent emphasis on identifying the drivers of sustainability, i.e. the major factors that have an influence on sustainable outcomes as well as the processes that precede them.

Dynamic Sustainability

Additionally, there is an increasing interest in "dynamic" sustainability, i.e. the achievement of structural change through evolution or innovation in response to changing external factors.[1] Dynamic sustainability implies adaptive (or developmental) approaches to planning, managing and/or evaluating interventions. Adaptation happens when a selection process leads to some improvement according to some measure of success, which has to make sense in the specific context. Adaptability in this sense is the capacity to recognize, understand and implement the adaptation required in a given situation. The "adaptive cycle" then is a model for understanding the recurring stages in which a community is more or less likely to respond to programs designed to promote sustained change. The benefit of understanding the adaptive cycle is therefore centered on understanding a system's resiliency, i.e. its capacity for coping with

[1] Interest in dynamic sustainability is confirmed by lessons learned from recent evaluations of USAID support to basic education.

change. When evaluating sustainability in this adaptive sense, issues to be considered notably include fidelity of implementation, thresholds or levels of sustainability, and the efficacy in reaching outcomes.

Dynamic sustainability is also closely linked with significant systems change and there is a growing interest in a systemic understanding of sustainability, which emphasizes the interaction between an intervention and its context (including wider policy systems). Several contributions from the systems field have been identified that can be used in the forthcoming USAID evaluation series on basic education. Some are located at the conceptual level (i.e. non-triviality, viability, feedback, complex adaptive systems), others at the level of methods and techniques (e.g. feedback loops, causal loop diagrams, stock-flow-diagrams). Other methods – notably those associated with the systems concepts 'perspectives' and 'boundaries' - would also be available (e.g. for identifying alternative framings or dealing with different perspectives) but would demand specialized skills and expertise.

Understanding Sustainable Outcomes

There are different ways of understanding the term "sustainable outcomes," yet there is a tendency towards pathway models, which foresee various possibilities to achieve sustainable outcomes and thus provide options for addressing sustainability in evaluations (e.g. at the level of activity, capacity, ideas) towards planned approaches to sustainability. Some attempts have been made – notably in the fields of health and social development – to conceptualize the processes for achieving sustainable outcomes in the form of frameworks that were developed essentially for planning and management purposes.

Evaluating Sustainable Outcomes

With respect to the evaluation of sustainable outcomes there are some lessons that can be used as a foundation; notably with respect to the role and focus of evaluations as well as tools for measurement and data collection. For the evaluation of sustainable outcomes (and the likely evaluation questions) several suitable design options have been identified: theory-based evaluations, participatory approaches, natural experiments and synthesis studies. Finally, several methods are recommended that fit with these design options and are in line with a dynamic as well as systemic understanding of sustainability: They include Realist Evaluation, Realist Synthesis and Contribution Analysis as well as two methods from USAID's Complexity Aware Monitoring initiative: Outcome Harvesting and Process Monitoring of Impacts.

The task ahead now is to reflect on the relevance of this rather broad menu for the specific aims and needs of the forthcoming evaluation series focusing on sustainable outcomes in basic education, which is to be carried out in a partnership arrangement between several teams/working groups within and outside USAID. Since there are no off-the-shelf methods available for this evaluation series, this reflection and the corresponding choices should be guided by creativity, aiming for mixes of methods or techniques that are appropriate for the specific situation (and thus might differ for project types or the time elapsed after an intervention). The fact that this evaluation series can make comparisons across programs and contexts provides a good opportunity for establishing a coherent and systematic knowledge base about program sustainability.

INTRODUCTION

USAID's Office of Learning, Evaluation and Research, in the Bureau for Policy, Planning and Learning (PPL/LER), is undertaking a series of ex-post evaluations (i.e., evaluations that take place several years after projects are completed) on 'sustainable outcomes' of USAID programs. The purpose of this evaluation series is to better understand what factors promote sustainable outcomes in development programs and projects. The series will take a systematic approach to understanding a variety of perspectives and intended and unanticipated outcomes. Basic education activities will be the focus for an initial round of evaluations and provide the context for site selection and determining activities and outcomes on which to focus as a starting point. The E3 Analytics and Evaluation Project, led by Management Systems International (MSI) along with partners Development & Training Services (dTS) and NORC at the University of Chicago, is supporting the design and implementation of this evaluation series.

This literature review is intended to support the evaluation project team by addressing the evidence or evidence gaps related to factors that contribute to the sustainability of development program outcomes, in particular with respect to systems and complexity theory and thinking.

Audience

The primary audience for this review is USAID staff involved in the design of this evaluation series, members of an Evaluation Advisory Group, the E3 Analytics and Evaluation Project team supporting the evaluation design process, and the evaluation team(s) that will conduct the evaluation series. This literature review will inform future decisions about the scope and parameters of the ex-post evaluation series. It is also anticipated that this literature review will serve as an ongoing resource for the research process over the course of the evaluation series.

METHODOLOGY

Part 1 of the review addresses the use and operationalization of the terms "sustainability" and "sustainable outcomes" and the relationship between sustainability and adaptability. It specifically addresses the following questions:

- What is sustainability?
- What is the relationship between sustainability and adaptability?
- What does the systems field contribute to this discourse?
- How might "sustainable outcomes" be defined?

Part 2 identifies lessons learned from previous evaluations and proposes some designs and methods that are appropriate for evaluating sustainable outcomes.

The literature review draws on a range of documents, including academic articles, working papers, discussion papers, book chapters, books case studies and relevant grey literature. The review was not intended to be a comprehensive search of all materials relating to the topics, but is based upon a curated selection of documents that will be most helpful for designing this evaluation within the boundaries of sustainability and systems. The review was also based on previous work that Mr. Richard Hummelbrunner, the review's main author, was involved in, notably the 'New Voices' Consultation on the preliminary evaluation questions for this evaluation series.

The criteria used for selecting the documents included in the review were whether the document addressed one or more of the research questions, was cited within a range of preselected documents on the subject already known to the reviewer, and covered the fields of either/or international development or social development (including basic education and health). The documents cited in section titled "Contributions from the 'Systems' field to the Sustainability Discourse" and "Appropriate Designs for Ex-Post Evaluations of Sustainable Outcomes" were already familiar to the reviewer based upon his professional background and experience, both as an evaluator and a practitioner in systems thinking.

This review does not investigate the rather extensive fields of sustainable development and sustainability science. Nor were the lessons learned from previous USAID evaluations in basic education captured in detail, as they are either already known to the evaluation team or will be the subject of further review to be carried out by them.

Annex A is a bibliography of the literature that was used in this review (in alphabetical order). References to this literature are made throughout in the text. In addition, a number of websites that were used in the review and that are considered particularly useful for the subject of the review are also included in Annex B.

FINDINGS

Part 1: Use and Understanding of the Terms "Sustainability" and "Sustainable Outcomes"

Discussions of sustainability have become an important part of international development discourse and various approaches to measuring and evaluating sustainability have been proposed. The diversity of these approaches can be confusing as sustainability has come to mean different things to different people and varies across contexts. It is therefore important in discussions about evaluating sustainability to clarify the term and then to operationalize it for the purpose at hand.

Daniela Schröter (2010) has identified two major tendencies in evaluating sustainability:

- Concern with macro-level issues of sustaining social and economic development under consideration of environmental protection. While this understanding prevails in the sustainable development field, it has evolved considerably and now focuses on not only environmental factors, but also embraces issues like participation and human rights.
- Concern with micro-level issues; meaning the continuation of interventions that receive time-limited funding. This understanding dominates in the North American evaluation literature, where sustainability is primarily thought of in terms of continuing program activity (and its benefits) beyond initial funding cycles, e.g. via diversification of funding or institutionalization.

It is this second type of understanding that forms the basis for the current literature review, as it corresponds with the understanding - and the needs - of the envisaged evaluation series.

Definition(s) of Sustainability

Sustainability is one of five OECD/DAC evaluation criteria and is defined as "measuring whether the benefits of an activity are likely to continue after donor funding has been withdrawn" (OECD 1991). When evaluating the sustainability of a program or a project, the OECD has recommended considering

the following questions:

- To what extent did the benefits of a program or project continue after donor funding ceased?
- What were the major factors that influenced the achievement or non-achievement of sustainability of the program or project?[2]

While the OECD/DAC criteria have remained relatively unchanged since they were first formulated in 1991, others have questioned the conceptualization of the sustainability as contained in the criteria. In a critical assessment, Chianca (2008) noted with respect to the sustainability criterion that the definition is limited to prospective sustainability and does not include retrospective sustainability. Furthermore, it only mentions the need to consider environmental and financial aspects of sustainability, leaving out other essential elements to the sustainability of interventions such as political support, cultural appropriateness, adequacy of technology and institutional capacity.

Among Chianca's proposals for improvement are the following:

- Broadening the definition by including aspects beyond financial and ecologic issues. He specifically cites the Danish International Development Agency (DANIDA), which has identified seven determinant factors in its definition for sustainability of aid interventions: policy support measures, choice of technology, environmental matters, socio-cultural aspects, institutional aspects, economic and financial aspects and external factors.
- Introduce "quality of process" and "exportability" either as new criteria or amendments to existing criteria. Aspects of process quality include coverage, responsiveness, stakeholder participation and cultural appropriateness. Exportability determines the extent to which an aid intervention – as a whole or partly – is transferable to another setting.
- Introduce bars, i.e. minimum acceptable levels of performance, below which an intervention will be considered fully unacceptable regardless of its performance on other evaluation criteria. This notably applies to negative side effects as a dimension of the impact criterion, i.e. if an aid intervention is affecting the people or the environment in any serious detrimental way. A bar should be placed on sustainability as well, but only after the intervention clears the bar in the other four criteria.

Better articulation of the criteria is one approach to improving the evaluation of sustainability of international development interventions, but proper application is another and perhaps even more important step. In a recent blog on evaluation, Deborah O'Connell lamented that the OECD does not set any standards for applying this criterion in an evaluation. She pointed out that even with widely accepted definitions for sustainability, operational tools will always require judgments in terms of "sustaining what, for whom, where, and for how long." These questions are laden with human values and social choices that are context-specific and differ across time, space and culture.

There is also a tendency to view sustainability as a series of processes that take place during the earlier stages of a project's life cycle, as well as by developments that take place in the project environment (Johnson et al. 2004, Pluye et al. 2004). This process definition of sustainability allows planning for and integrating sustainability in ongoing activities by identifying and monitoring a range of influence factors. This implies that thinking about sustainability must start early and that program staff and donors can enhance these factors during implementation in order to increase the probability that program outcomes will be sustained.

[2] It is also postulated that projects need to be environmentally as well as financially sustainable.

Marie Ann Scheirer and James Dearing (2011) have pointed out that studies tend to examine different factors or rely on informants to suggest what they believe influenced sustainability instead of systematically assessing any common set of predictive factors. They also point out that the process definition of sustainability presents specific challenges for evaluation, as it requires the explicit definition of outcome variables and measures of hypothesized influences on those outcomes, They propose a common set of process factors as predictors of sustainable outcomes in public health programs, which are conceived as variables (dependent and independent) and assembled in form of a sustainability framework to guide monitoring and evaluation (see below).

With respect to assistance to education, Gillies (2010) has noted that sustainability is usually assessed against one of two criteria: continuity of project activities and financial responsibility. However, the definition of sustainability as a scheduled exchange of funding responsibility from donor to recipient country has been revealed to be impractical and even counterproductive. Aid 'reforms' conceived as efforts in speeding up the transfer of financial responsibility led to premature interruptions in programs, which caused them to be even less effective and sustainable. Gillies therefore suggests defining sustainability in a developmental sense – programs would be focused on initiating and stimulating change in the education system and less concerned with initiating activities that cannot be continued without on-going subsidies.

This last definitional proposal echoes the discussion in other sectors and is related to fundamentally different ways of conceiving sustainability. With respect to enterprise development and based on an analysis of evaluations of the 'Making Markets Work for the Poor (M4P)' approach, Tim Ruffer and Elise Wach (2013) have identified two conceptual models of sustainability:

- Static sustainability is defined as the extent to which the status quo (in terms of the results achieved through an intervention) will be maintained after external support is withdrawn. In other words, the extent to which the project's legacy is maintained. This concept is about maintaining the status quo achieved by an intervention through the replication of activities, processes or outcomes.

- Dynamic sustainability is defined as the achievement of structural change that enhances resilience to shocks and stressors through evolution or innovation in response to changing external factors. Dynamic sustainability is inherently linked with significant systems change.

They conclude that while sustainability should be defined in terms of both static and dynamic sustainability, only the latter will take full account of the extent to which changes become embedded in a wider system (in this case the market). This dynamic dimension to sustainability suggests that underpinning the outward or superficial performance of any 'sustainable' system are a variety of critical but often less visible institutions and functions. This includes the resilience of the system to shocks and stress and the capacity to evolve or innovate in response to a changing external environment.

In an evaluation of the sustainability of German development aid for professional training, Rheinhard Stockmann (1993) proposed an extended version of dynamic sustainability based on understanding interventions as being double embedded within an implementing organization and a wider implementation context. Therefore interventions can unfold their effects within and via their implementing organization, which is also the filter or transmitter of contextual influences affecting the intervention. Effects are sustainable if they display one of the following features:

- The processes and structures established by the intervention enable the implementing organization and/or beneficiaries to make structural adjustments in case of changes in the implementation context (problem solving function).

- The benefits of an intervention extend beyond the beneficiaries and/or implementing organization, thus the outreach of an intervention is improved and more people can benefit (multiplier function).

- The processes and structures established by the intervention become a model for others, thus copying successful patterns for appropriate problem solving capacity (model function).

Sustainability and Adaptation/Adaptability

The dynamic understanding of sustainability discussed above is based on the concept of adaptation, which is the foundation of evolutionary processes. Robert Axelrod and Michael Cohen (2001) suggest that adaptation happens when a selection process leads to some improvement according to some measure of success. The success is measured in a way that makes sense in the specific context. In natural evolution, success is measured as the fitness of an organism, in the economy it might be measured in profit produced by a specific business model.

Whereas adaptation evolves new traits, the shift in the function of a trait is called exaptation. This happens when an existing capability is used for another purpose. An example from biology are feathers, which originally were developed for temperature regulation but were later also used for flight. But exaptation does not happen only in natural evolution, but also by intention. In other words, intentional exaptation occurs when a tool is designed for a specific task and is then 'co-opted' to perform another task that is not related with the original intent. In a recent blog, David Snowden (2012) referred to exaptation as "serendipitous innovation" and cites the example of the development of Viagra that was originally intended as treatment for angina pectoris.

In his discussion of innovation, Max McKeown (2012) has defined the adaptability of social groups or individuals using a three-step model called RUN: (1) recognition of required adaptation; (2) understanding of adaptation required; and (3) necessary adaptation. He sets out four levels of adaptation that social groups can achieve:

- "Collapsing" is the end of the social group.

- "Coping" is survival "without prosperity, pride or joy."

- "Thriving" provides benefits that are worthwhile and desirable.

- "Transcending" overcomes the limits of the old system, allowing more for everyone.

McKeown believes that usually a mix of different levels of adaptation can be claimed. Many efforts to reach higher levels of adaptation fail because one or more of the steps (i.e., RUN) is not completed successfully: people may fail to understand change is necessary, or not understand what kind of change is necessary, or simply avoid making those changes.

Based on research of ecosystems dynamics, Gunderson and Holling (2002) have developed a model of adaptation known as the "Adaptive Cycle." It takes the form of an infinity figure constructed along two dimensions of a matrix: the horizontal dimension expresses diversity along a connectedness continuum, and the vertical dimension expresses the potential measured by the degree of release (or storage) of resources. These two dimensions form a matrix of four quadrants, which represent stages in a recurring adaptive cycle (see Figure 1).

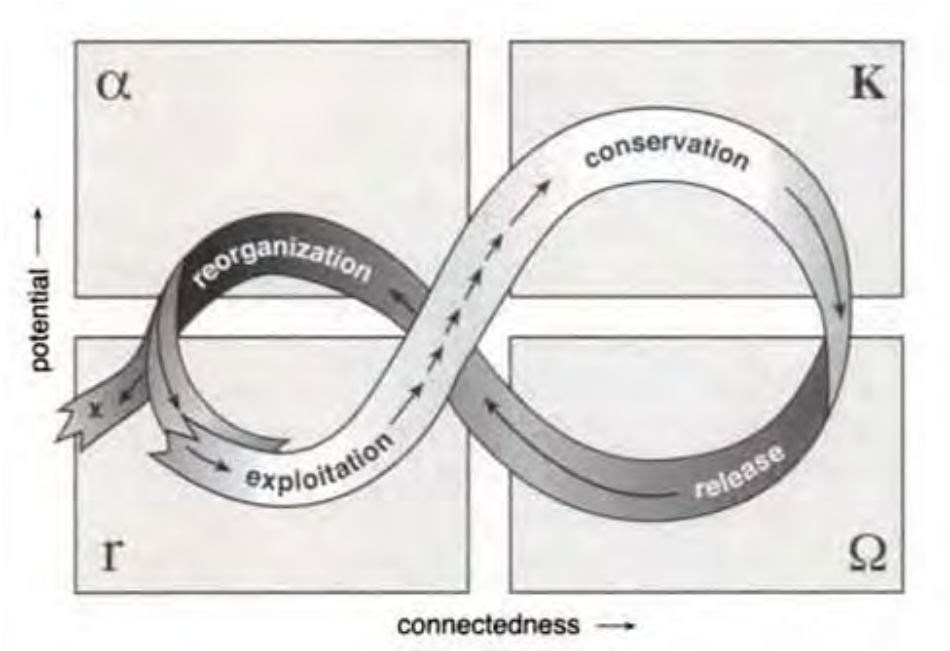
- **Release:** Resources are set loose due to some act of destruction (e.g. forest fire), ending the previous dominance of some species

- **Reorganization:** the destruction during the release phase opens the way for new beginnings and new growth

- **Exploitation:** accelerated growth of some varieties over others in the competition for resources

- **Conservation:** Diversity is relatively low and resources are stored in few dominating species

FIGURE 1: ADAPTIVE CYCLE MODEL AS DESCRIBED BY GUNDERSON AND HOLLING (2002)



The significance of the adaptive cycle with respect to sustainability lies in understanding the relationships and time flows among the quadrants. Ecosystems slowly progress from exploitation to conservation by increasing their connectedness and stability, accumulating a capital of resources. The transition from conservation to release, reorganization and then back to exploitation can take place rather quickly. However, these transitions are not guaranteed and ecosystems can become trapped in one phase, which threatens their sustainability and long-term existence.

The adaptive cycle cannot only be observed in nature, but also in social systems like communities, organizations or programs. Westley and Miller (2003) have applied it to innovation and change processes and have identified different 'psycho-social' regimes that can be associated with each of the quadrants:

- Release is characterized by "creative destruction" when resources are lost and stress and confusion are building up.

- However, those who thrive on crisis and see the potential for new opportunities facilitate the transition to reorganization characterized by "exploration," i.e. widespread experimentation, innovation and renewal.

- Exploitation is marked by "selection," which leads to making choices and focusing attention on the most promising ideas.

- As winners or best practices become dominant, competition is reduced which leads into the conservation phase dominated by a "bureaucratic" mindset. This favors stability, standardization and the increase of efficiency or scale, which in turn increase rigidity, inflexibility or the resistance to new change.

The adaptive cycle is based on a new understanding of stability exemplified by the term resilience, i.e. the ability to cope with change (Gunderson and Holling, 2002). Traditional forms of "engineering resilience" focus on maintaining the efficiency of function based on minimizing threats to performance and stability near an equilibrium steady state. Sustainability is measured by the resistance to disturbance and speed of return to the equilibrium. Ecosystem resilience on the other hand focuses on maintaining the existence of function based on adaptability and responsiveness. Sustainability is measured by the magnitude of disturbance that can be absorbed before the system changes its structure by changing the variables and processes that control behavior. A system is resilient when it can modify its structure and is able to adapt and evolve, i.e. developing new solutions for new problems it faces.

In recent years resilience has increasing been used as a heading for thinking about sustainability in international development, e.g. in food security, adaptation to climate change, linking relief efforts to development or addressing the vulnerability of individuals or communities. Levine (2014) critically examines current approaches to developing measures for resilience and cautions against the sole reliance on quantification as well as against tendencies to 'reify' resilience, i.e. to conceive "resilience-building" as a separate group of activities. Instead of the search for universal measures of resilience, he proposes to improve methods for quantifying the things that really matter. This implies that instead of developing new methodologies for quantifying resilience, existing good practice around analysis, assessment and monitoring should be used more frequently to address the needs quantification is intended to meet.

Scheirer and Dearing (2011) have pointed out two issues that arise when sustainability in the adaptive sense as outlined above is to be evaluated:

- Fidelity of implementation: How much adaptation or change of the program components can occur while still defining the intervention as "sustained"? If modifications have been made, is it still the same program? This would require defining the core components and customizable components of an intervention, in order to guide both implementation and sustainability.

- Threshold: what extent or components of the program must be present for a program to be counted as sustained, especially if the program is implemented across a number of sites? This is a dosage or induction strength question. Defining levels of sustainability with explicitly described sets of components for each level can reduce the need to dichotomize an outcome variable into sites that did or did not sustain an intervention.

Last but not least, a dynamic understanding of sustainability also needs to consider the issue of efficacy. In principle, only those interventions should be sustained that can providence evidence for efficacy in reaching their stated outcomes. Gillies has emphasized this point with respect to assistance to education when he states:

> "Activities should not be sustained simply because they are initiated— they should be sustained if they have demonstrated measurable, cost-effective and meaningful improvements over alternative approaches. Nor should they be sustained *ad infinitum*. The process of development is inherently a process in which interventions and activities change and improve rather than remain static at an achieved state. The concept of development itself argues against a narrow or static vision of sustainability" (Gillies, 2010).

In practice, the question of efficacy may be ambiguous: intended outcomes may only have been achieved for parts of a program (or at some sites) or only some of these outcomes show statistical significance. In addition, for some interventions evidence on efficacy may only accumulate over time, which poses issues of timing. The achievement of sustainable outcomes will likely require processes that start long before the initial funding ends – and evaluation results on sustainability might not be available when decisions about continuing a program have to be made. In this case, the problem of prematurely abandoning potentially useful interventions becomes salient.

This dilemma of continuing interventions when the efficacy of the program itself is still uncertain has no easy solution. Scheirer and Dearing (2011) propose that in such a case intervention research and sustainability research should proceed in parallel. Preliminary findings (on the effects of an intervention) should be made available to program managers before the end of the funded period, to let them know whether promoting sustainability is worthwhile. Program managers should continuously observe whether intended outcomes are achieved for their interventions and should be attuned to new research findings from studies of the effectiveness of those interventions. Hence they advocate building in fast feedback cycles and formative or developmental evaluation practices during program implementation.

When summing up these various findings, one can conclude that there are adequate definitions for adaptation and adaptability as well as a model for the evolvement of adaptation, which is also applicable for social systems. This 'adaptive cycle' model is based on a new understanding of sustainability expressed by the term "ecosystem resilience." The focus is on maintaining functions through adaptation and where sustainability is measured by the magnitude of the disturbance that can be absorbed before structural changes are needed. When evaluating sustainability in this adaptive sense, issues to be considered notably include fidelity of implementation, thresholds or levels of sustainability, and the efficacy in reaching outcomes.

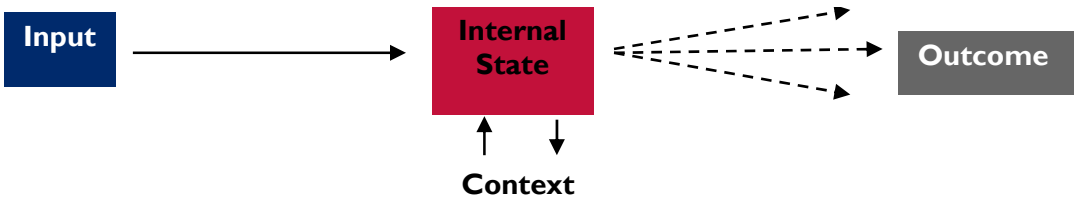
Contributions from the "Systems" Field to the Sustainability Discourse

The systems field is not a uniform stock of knowledge, but stems from research originally undertaken in diverse fields such as anthropology, biology, communications theory, sociology and management, which have mutually influenced each other. Therefore a single, concise and generally agreed definition of a system does not exist. Moreover the field is rather large and diverse concerning theories, approaches or methodologies. This section presents some selected elements from the systems field and briefly outlines their implication for the sustainability discourse.

Conceptual Models of Systems

Perhaps the earliest contributions with respect to a systemic understanding of sustainability come from the cybernetics tradition, i.e. the study of control and communication in mechanics and organic life. Heinz v. Foerster (1961) distinguished between two types of machines respective behavior: trivial machines have a simple mechanism whereby the same input will always lead to identical output, regardless of circumstance. **Non-trivial** machines can react differently to the same input – depending on their internal state and the context in which they are embedded (see Figure 2).

FIGURE 2: PROCESS BY WHICH NON-TRIVIAL MACHINES REACT TO INPUTS



Foerster cautioned against what he called the "mental trivialization" of social systems, i.e. regarding them as trivial (instead of non-trivial) machines. Human cognition has the tendency to think in simple cause-effect patterns and our everyday thinking functions along linear rules. Therefore the most frequent reaction when faced with complexity is to either ignore it or reduce it mentally in order to cope with it. But such reductionism strips social systems of their most important qualities, i.e. their internal dynamics and self-organization. Hence, he advocated replacing mechanistic thinking by systems thinking, which can help to avoid such inappropriate simplifications.

Interventions like development aid programs can be understood as non-trivial entities. Effects and their sustainability can neither be explained from inputs nor the internal state (of the structure transforming these inputs) alone, but from their interaction within a specific context. Such interventions will neither produce their effects in a linear way, nor will the effects be limited to those intended at the design stage. In order to assess the effectiveness and sustainability of interventions one should look at the entire range of processes triggered and outcomes produced, irrespective of whether they are in line with original intentions.

Moreover, differences in the achievement of effects can be used to assess the appropriateness of an intervention's logic (or theory) in view of the given operational context. These can provide indications about the internal dynamics and self-organizing forces that are at work within the implementation system or the operational context. Therefore exceptions, discontinuities, unexpected results and surprises in the data are valuable sources of information that can provide useful clues (e.g. for relevant internal/external changes, newly emerging challenges, novel ways of dealing with problems) that can help to better understand the likeliness and the conditions for producing sustainable effects.

Another important contribution from cybernetics is improving the understanding and operationalization of **viability**. In the 1970's, Stafford Beer explored the extent to which the viability conditions of living systems, such as biological organisms, relate to socio-cultural systems, such as organizations and enterprises. On this basis, he identified a number of requirements that must be in place if a social system is to prove viable, which he assembled in the Viable System Model (Beer, 1979). This model illustrates the minimum requirements that must be in place if social systems are to provide enduring results and be capable of further development. It is an ideal model for attaining optimum performance and for sustaining it in the long term, against which the actual state of an organization can be assessed.

Viable System Model (VSM) divides systems into five sub-systems interacting at three levels:

- The **operational level** is made up of those units that actually do the work and are responsible for producing outputs.
- The **strategic level** consists of three sub-systems that hold the operational units together: coordination, environmental monitoring, management and control.
- The **normative level** establishes policy in light of competing demands between the present and future as well as between internal and external perspectives.

VSM specifies how those sub-systems are to be connected, i.e. the structure and process of information that should flow through an organization so it can be viable, and identifies what kinds of decisions need to be made by whom and on the basis of which information.

Feedback Loop Conceptualizations

When elements of a system are connected, every action can ultimately also have repercussions for the actor itself. Thus "linear" cause-effect patterns are replaced by reciprocal "circular" interaction patterns,

where actions can be both cause and effect. Cybernetics has developed a language for representing circular relations based on the concept of "feedback." **Feedback** is the transmission or return of information and a feedback loop is a closed sequence of these transmissions between various elements. Feedback loops appear in two types:

- **Reinforcing** or positive feedback refers to a situation where elements respond to perturbation in the same direction (as the perturbation), i.e. "A produces more of B which in turn produces more of A". Change in one direction is compounded with more of the same change (sometimes referred to as 'cumulative causation'), which can produce growth or decline, depending on the direction of change.
    - A savings account is an example of positive feedback, as the amount of savings and the interest earned are linked in a reinforcing manner: If savings increase they will lead to higher interest, which again are added to the savings and so forth. Contrary, if savings decrease the interest earned will also go down which again affects accumulated savings.
- **Balancing** or negative feedback occurs when elements respond to perturbation in the opposite direction, i.e. "A produces less of B which in turn produces more of A". Thus changes are attenuated and the system will tend to be stable by closing the gap between a desired and the actual state.
    - An everyday example of negative feedback is a thermostat: A particular room temperature is set as the heating system goal. When the temperature in the room reaches an upper limit the heater is lowered so that the temperature begins to fall. When the temperature drops to a lower limit, the heater is turned up again. Provided the limits are close to each other, a steady room temperature is maintained. This example also illustrates the concept of goal setting, because it is the gap between the set point and the actual temperature that drives and stabilizes the heating system.

Feedback loops can be used for articulating the dynamics of a given situation and are associated with specific patterns of behavior: Reinforcing feedback leads to exponential growth (or decline) and escalation, whereas balancing feedback results in stabilization or goals seeking behavior, where a variable starts either above or below a goal level and over time moves towards a goal or platform. Reinforcing feedback by itself is unstable, amplifies small effects and tends to get out of control. But it is usually counteracted by a balancing feedback of some sort (e.g. spending money which slows down the growth of a savings account). Another important piece of information for understanding feedback dynamics is the quality of the relationships between elements: For example the degree of influence exerted (e.g. strong, medium, weak), its temporal duration (e.g. short-, medium-, long-term), the continuity of an influence over time (e.g. delays or time lag), or non-linear sequences (e.g. exponential curves, 'thresholds' or 'tipping-points').

Feedback loops are the building blocks of **Causal Loop Diagrams** (CLDs) that visualize the relationships of elements over time through the interaction of these two types of feedback loops. CLDs have been widely applied since the 1970's in diverse fields, including environmental management and sustainable development. CLDs give a quick synoptic overview, can be modified rather easily and facilitate communication on complex issues, thus they are appropriate for collaborative settings. CLDs permit analyzing current states or relational patterns, but also making assumptions about future behavior, e.g. on the likeliness for sustaining outcomes.

CLDs are especially appropriate for understanding generative mechanisms, i.e. the processes that bring forth observable phenomena like outputs or outcomes. They allow for a systemic understanding of situations by mapping the structure that is responsible for producing patterns that recur over time.

Therefore CLDs can be particularly useful when dealing with situations that are influenced by numerous and interrelated factors, which are difficult to structure verbally. By visualizing elements and their linkages one can understand how a change in one element might lead to changes in the whole situation. This allows one to understand the factors necessary to sustain a certain pattern – or identify suitable points of leverage or entry that can be manipulated for sustaining a pattern (Meadows 2008). Feedback loops and CLDs can also be used in combination with frameworks for sustainability (e.g. for modeling influencing factors for sustainability) or in cases where pathway models for sustainable outcomes are applied (see section below).

Complex Systems Models

CLDs are useful for clearly defined issues and relatively closed settings that lend themselves to simplifying diagrams. They are less appropriate for larger scale processes. Predicting a system's behavior is only possible in the case of simple CLDs, but this becomes very difficult once feedback loops are interacting and delays are introduced (Richardson, 1986). To predict patterns of behavior the computer-based models of **Systems Dynamics** are more appropriate as they allow simulations. They are based on the idea that systems consist of elements that have, at a specific point in time, a value ("stock"), which can change over time through inflows or outflows. For example, in a bank account savings are the accumulating (stock) variable, whereas interest and spending are in and outflows. The dynamic behavior of a situation is explained by the relationship between stock and flow variables, which is expressed in stock-flow diagrams (Forrester 1961, Sterman 2000). Stock and flow variables must be quantified and their relationships defined in mathematical terms. However, these quantities don't have to be real numbers and can be symbolic.

Systems Dynamics offers valuable insights into the dynamic behavior of complex systems and has been widely applied with substantial success. It has been very influential for the sustainable development movement, in particular through the models developed by Dennis and Donella Meadows (1972) that explored how exponential growth interacts with finite resources. Systems Dynamics models require some expertise in modeling and mathematical thinking, but since the 1980s easy-to-use software is available for creating them. They can be used before (e.g. for testing hypothesis on the achievement of outcomes) or after a program is implemented (e.g. to understand why outcomes were not achieved as planned). Forrester (1993) emphasizes that they can provide understanding about puzzling or controversial data. Frequent dangers are the dominance of experts and regarding the models as representations of reality. To curb this risk, stakeholders should be involved in the modeling process and the reflection on findings. In this way the models are used primarily as a device for aiding communication between stakeholders and for the development of deeper understanding on complex issues. The expertise required for this lies rather more in facilitation than in modeling.

Complexity science has developed as an approach to study situations of "organized complexity," where many parts interact and produce behavior that cannot be explained in terms of interactions between the individual elements alone. It provides a diverse set of tools and approaches, which are of particular relevance for issues in international development (Ramalingam 2013). One of the best known is the idea of **Complex Adaptive Systems (CAS)** that consists of semi-autonomous agents who interact to form system-wide patterns. CAS are characterized by two distinct principles:

- **Emergence**, a process whereby larger entities or patterns arise through interactions among entities that themselves do not exhibit such properties; and
- **Self–organization** among these entities, due to the absence of a central command structure.

CAS are open to their environment, thus systems and their context mutually influence each other, continuously adapt and co-evolve.

The CAS concept has taken the notion of non-trivial behavior further and allows for operationalizing it. The behavior of a CAS is determined by internal structure as well as contextual relations. This means for instance that external agents can attempt to influence the achievement of effects from the outside, but these influences are ambivalent, as they can disturb or stimulate further development. Moreover, contextual influences are modified and rearranged in line with a system's internal structure and logic. This implies two things with respect to outcomes and their sustainability:

- While the emerging patterns of effects can be observed, they can neither be predicted nor do they depend on the starting conditions alone.
- Patterns of effects are not under the control of external actors like aid agencies. Sustaining outcomes in particular depends very much on local stakeholders – which is also an expression of the principle of local ownership.

Three Concepts Inherent in Systems

In the historical development of the systems field three **Systems Concepts** have emerged as the essence of systems thinking (Midgley 2000). Williams and Imam (2007) have explored their relevance for the evaluation field, for which these concepts are not new. But the systems field can offer specific contributions in terms of characteristic aspects and methods that underpin them:

- **Interrelationships:** This is probably the most familiar concept, because it is the oldest and the one most strongly embedded in the popular use of the term system (e.g. education system, health system). It essentially means that a set of elements or actors is somehow interconnected. The systems field can particularly help in analyzing dynamic aspects (i.e. when interrelationships affect the behavior of a situation over time), non-linear aspects (i.e. where the scale of effect is apparently unrelated to the scale of the cause) and the sensitivity of inter-relationships to context (i.e. where the same intervention in different situations or areas leads to different results). Most of the visualizing and modeling methods described above are associated with this concept.
- **Perspectives:** This concept implies that the same situation can be "seen" in different ways, which can generate better insights into the actual behavior of programs, because people usually behave on the basis of their perceptions of what is, or what should be, rather than some official logic defined by someone else. The systems field has a specific take on these framings by seeing them as a combination of stakeholders (people or things that have a common role in a situation) and stakes (related to individual values and motivations). Thus, different stakeholders may share the same stakes, and a stakeholder group might contain several different (perhaps conflicting) stakes.
- **Boundaries:** Every endeavor has to set boundaries that define what is included and what not, what is considered as (ir-) relevant – and by whom. Boundaries are fundamentally about values and power issues, as it is the dominant perspective(s) that decides the boundaries. Key boundary decisions of an intervention include purpose and outcomes, intended beneficiaries, measurement approaches, decision-making authority or expertise. Applying this concept means to set boundaries consciously and to critically consider the implications of boundary decisions with respect to concerns of ethics, legitimacy and marginalization.

The systems field provides a range of methods that can be associated with each of these concepts (Williams and Hummelbrunner 2011), but only a few could be outlined in the framework of this literature review. With respect to the evaluation of sustainable outcomes, methods for identifying alternative framings, dealing constructively with different perspectives or for reflecting on boundaries might be particularly relevant.

Possible Definitions of "Sustainable Outcomes"

In their meta-evaluation of USAID assistance to basic education, Chapman and Quijada (2009) found that no substantive conclusion can be made about program sustainability. Although sustainability was a stated goal and criterion of success in many of the programs, they have generally failed to either define or measure sustainability. The difficulty in defining sustainable outcomes partly stems from the diversity of views about sustainability – and is not unique to USAID. In a review of lessons learned over decades of foreign aid to education, Abby Riddell (2012) found little evidence for the sustainability of aid interventions and quite diverse approaches for assessing it. She concludes that for aid to make lasting improvements a more systemic approach is needed that focuses on building improved education systems and pays more attention to the interface of aid with these education systems as well as context in general.

A clear and evidence-based definition of sustainable outcomes could also not be found in other recent meta-studies of aid in basic education (European Court of Auditors 2011, Glewwe and Kremer 2005). Therefore it seems appropriate to look beyond the sector and review how this issue has been dealt with in other areas of social change, in particular the health and social services sectors, where sustainability has received considerable attention.

While understandings of the term "sustainable outcomes" are diverse, efforts have been made to conceptualize it. Shediac-Rizkallah and Bone (1998) in their review of health promotion literature identified six different ways in which the term sustainability was used, which also illustrate different meaning of the term outcome:

- The capacity to service (i.e. support) coverage at a level that will provide continuing control of a health problem.
- The capacity of a project to continue to deliver its intended benefits over a long period of time.
- The ability of a program to deliver an appropriate level of benefits for an extended period of time after major assistance from an external donor is terminated.
- The long-term viability and integration of a new program within an organization.
- The process by which new practices become "standard business" in an organization (i.e. their routinization, institutionalization or incorporation into an organization).
- The development of the capacity (knowledge, skills and resources) of the organization to conduct effective programs.

In a more recent review of empirical studies of the sustainability of health-related programs in the US and Canada, Scheirer (2005) has confirmed this list but identified two additional types of sustainability:

- Sustaining the "ideas, beliefs, principles or values underlying the initiative."
- The sustainability of the implementing organizations themselves.

In an evaluation of the sustainability of services for young children, Rogers and Williams (2008) have built on these definitions and proposed a pathway model for assessing sustainable outcomes. They found that although sustained results can be directly achieved through a project, more commonly they are achieved through pathways that involve one or more of the following:

- Sustained activities by organizations – these might be the same activities undertaken during the project or different activities.

- Sustained capacities - either of the families or of the organizations – to provide benefits for the children and achieve the results.

- Sustained idea or service model – including general approaches to working with families and specific programs.

In a personal communication, Bob Williams has outlined how this pathway model relates to different dimensions of sustainability:

- Understanding how to sustain a particular activity enables that activity to be transferred from one setting to another.

- Understanding how to sustain outcomes enables the activity to be scaled from pilot project to mainstream.

- Understanding how to sustain a fundamental idea or model enables adaptability in more complex and diverse situations.

In order to improve the achievement of sustainable outcomes in health programs, it was considered helpful to move away from seeing sustainability as a latent goal towards a more planned approach, i.e. active attempts to maximize the potential for sustainability. This foremost requires a better understanding of the conditions and influencing factors for sustainability. Shediac-Rizkallah and Bone (1998) proposed one of the first frameworks for conceptualizing sustainability, in this case for community-based health programs. They identified three categories of influencing factors, which can function as guidelines for sustainability planning:

- Design and implementation factors, e.g. type of program, negotiation process, duration, financing, effectiveness and training component of the program.

- Factors within the organizational setting, e.g. institutional strength, integration with existing services, institutional support.

- Factors in the broader community environment, e.g. community participation, socio-economic and political considerations.

Over the years a number of frameworks for sustainability have been developed in the health and social services sectors, but only a few can be mentioned here:

- In her review, Scheirer (2005) has concluded that for public health programs an overall framework of influences on sustainability is emerging, and she categorized the factors under three headings: (1) Characteristics of the intervention, (2) Factors in the organizational setting and (3) Factors in the community environment of each intervention site. Scheirer and Dearing (2011) later proposed a more detailed conceptualization, which also includes financial resources and other inputs and considers all of these factors to be independent variables. Sustainability outcomes are seen as dependent variables. The framework summarizes the hypothesized relationships among the dependent and independent variables, and the degree to which they are embedded in a policy and financial context.

- For community-based family support programs, Mancini and Marek (2004) have proposed a framework consisting of seven major elements: leadership competence, effective collaboration,

understanding the community, demonstrating program results, strategic funding, staff involvement and integration, and program responsivity (understood as adaptability). They have also tested this framework empirically and found that leadership competence (which includes securing funding, quality control and evaluation) is the most important factor, followed by effective collaboration and program adaptability.

- For the education sector Nkansa and Chapman (2006) have proposed a framework for sustaining community participation in education programs that comprises eight managerial and socio-cultural factors: Planning, transparency, leadership, participation, social cohesion, resources, community kills and valuing of education.

The health sector also is a forerunner in the international development field with respect to the uptake of systems ideas, driven by an understanding that strong health systems are needed to improve health outcomes (Savigny and Adam, 2009; Sarriot, and Kouletio, 2014). This also contributed to a more systemic understanding of sustainability, whereby sustainable outcomes are seen as resulting from processes taking place in a local system where a wide array of stakeholders share responsibility to generate and maintain positive health outcomes.

A group of NGOs have partnered with USAID to develop a Sustainability Framework as a tool for planning and measuring community-oriented health programs (Sarriot et al, 2008). This Sustainability Framework offers an interactive model for assessing progress on critical dimensions: the health outcomes being promoted, characteristics of health services (quality, accessibility, equity), institutional capacity and viability of local government and civil society agencies with long-term responsibility for the outcomes, capacity in beneficiary communities (e.g., social capital, community organization, knowledge/skills, resource mobilization) and socio-ecological conditions enabling the work of these local agents. Based on a long-term application of this Framework in Bangladesh, Sarriot (2014) found that the Framework's most important contributions were as a tool for engagement of and negotiation between local stakeholders and for reducing the tension in balancing sustainability and equity.

This look into the health and social development sectors has shown a clear tendency towards planned approaches to sustainability and pathway models for defining sustainable outcomes. These foresee various possibilities to achieve sustainable outcomes and provide options for addressing sustainability in evaluations (e.g. at the level of activity, capacity, ideas). Several attempts have been made to conceptualize the dynamics of sustainability in the form of frameworks and due to a growing interest in a systemic understanding of sustainability some of the most recent frameworks are explicitly based on systems ideas.

Part 2: Evaluating Sustainable Outcomes

Lessons Learned About Evaluating Sustainable Outcomes

A first and major task in any evaluation is to determine focus and criteria. With respect to evaluation objects, Schröter (2010) distinguishes between evaluation *for* sustainability (contribution to sustainable development efforts) and evaluation *of* sustainability. This means determining the merit, worth and significance of efforts to continue a given program beyond the removal of initial program resources. She proposes a Sustainability Evaluation Checklist (SEC) for planning and designing evaluations of sustainability within development contexts. In addition to some general consideration (e.g. clarifying perspectives on sustainability) this checklist notably includes extensive sections on the criteria of merit, worth, and significance, that were developed based on an extensive literature review and feedback from sustainability evaluation experts and practitioners. Because the checklist is relatively generic, it may be

valuable in a heuristic sense for discussions within organizations and among evaluators, clients, stakeholders and those impacted by interventions.

With respect to evaluation's role in supporting sustainability, Weiss et. al. (2002) argue that evaluations can support sustainability either through strategy (by advising and facilitating strategy development) or by treating sustainability as a variable to be operationalized and tracked over time, thereby encouraging learning from an initiative's early stages. They consider the latter to be a more purposeful approach but found that in the field their research has covered (family support initiatives) few evaluators have taken this route. In order to operationalize sustainability they suggest data should be collected in four areas and explain how data should be examined through the lens of sustainability:

- **Organizations and/or Projects** (e.g. success in obtaining additional funding, presence of revenue-generating strategies or of multiple funders)
- **Ideas** (i.e. maintaining and diffusing the initiative's core principles, values or beliefs and commitment to continue work started)
- **Relationships** (e.g. continuous forms of collaboration, joint projects or products, collaboration that is not initiative-driven)
- **Outcomes** (i.e., codification of outcomes, support/demand for outcomes, continued involvement/commitment to outcomes over time).

A report to the World Economic Forum (O'Connell et al. 2013) points out that data and measurement are crucial but insufficient for enhancing sustainable outcomes, unless they are understood as part of a wider iterative process of learning and acting, and embedded in institutions that can provide the mechanisms for implementation. Based on a review of the multiple approaches and methods to sustainability, this report has identified four universal building blocks:

- Institutional mechanisms - the formal and informal rules that provide the governance, oversight and stability necessary to implement sustainability.
- Data – specification, collection, analysis and the use of projections: data, which can consist of measurements, modeled interpolations or projections, are used as the basis for evaluation.
- Evaluation - interpreting the meaning or value of the data in relation to agreed sustainability objectives.
- Feedbacks – the flow of information or action between components of the framework, including catalyzing changes that promote sustainability.

Greater availability of data will only be effective for sustainability if it is used within such an approach, and the report proposes a framework that outlines the linkages between these elements across a range of scales (e.g. local, regional, national).

In their meta-evaluation of USAID assistance to basic education, Chapman and Quijada (2009) have found that availability of local funding to pick up costs was the largest (but not the only) factor inhibiting continuation of these projects. Management capacity of government and lack of clear ownership of project activities were also cited.

Scheirer and Dearing (2011) have confirmed that these findings are also valid for public health programs. The availability of funding is a major influence for sustaining outcomes, other factors relate to characteristics of the intervention, organizational setting and community environment. They also found

that the sustainability literature has not developed any consensus or guidance on which type of sustainability outcome should receive priority. Often, the type of data collected depends primarily on the funding available. Collecting data about continued benefits for individual clients is usually more expensive than a survey of program directors. If feasible, the collection and assessment of information about long-term unintended consequences, including negative consequences of a program, could enrich the program sustainability knowledge base. Detecting these additional consequences may require continuing collection of qualitative data drawn from the broader context surrounding a specific program and its effects on a population.

A systemic understanding of sustainability, i.e. seeing sustainable outcomes as the result of a local system and its capacity to adapt and change, poses the challenge of evaluating systems and systemic change as they are related. Such an understanding is gaining recognition in the economic development sector, in particular in the field of 'Inclusive Market Development'. A recent literature review for the USAID sponsored Leveraging Economic Opportunities Project (LEO) provides an overview of various definitions of systems change as well as of frameworks for its evaluation. It also contains a section on monitoring systems change and a comparative analysis of indicators that are being used for this purpose.

In the public health sector there has been a debate about some fundamental questions with respect to sustainability, as exemplified by the following publications:

- Based on evidence collected from randomized evaluations, Kremer and Miguel (2004) have challenged the move towards sustainability, which stresses community mobilization, education, and cost-recovery. They state "teaching people to fish rather than providing fish is great if it works, but it works only if the donor knows more about fishing in the local area than the people who live there, and if the donor can transfer this knowledge". They concluded that while sustainability is certainly a desirable goal, it may be difficult to achieve and the pursuit of sustainability may be an illusion. They propose two policy alternatives as a way forward:
  - Donors should consider sustainability not on a project-by-project basis, but rather at the level of increasing overall national income.
  - Instead of creating an illusion of sustainability, donors should rather accept the reality of the need for continued subsidies for development projects and ensure that the necessary stream of funds for maintenance is available.
- Scheirer has concluded in her review titled 'Is Sustainability Possible?' (2005) that the evidence supported the possibility of project sustainability, under the right conditions generated by the convergence of internal and external supporting factors. The studies reviewed showed substantial convergence on five important factors influencing the extent of sustainability:
  - A program can be modified over time;
  - A "champion" is present;
  - A program "fits" with its organization's mission and procedures;
  - Benefits to staff members and/or clients are readily perceived; and
  - Stakeholders in other organizations provide support.

  However, she also noted that findings might have been quite different if different definition and measurement of the key outcome variable of sustainability or more rigorous methods of data collection and analysis had been used.

- Scheirer and Dearing (2011) raise the question whether sustainability is desirable and found that stakeholders have mixed opinions on this: there is support when an effective public health program has been implemented and should be sustained. But evidence on the sustainability of previous programs is readily dismissed when adoption and implementation is sought for new programs. They also caution that sustaining a program within an ongoing organization could become a hollow shell of activities perpetuated for their own sakes, especially if benefits for clients are not achieved. Instead of seeking institutionalization, it is often the capacity building and innovativeness generated by a program that should sustained.

Appropriate Designs for Ex-Post Evaluations of Sustainable Outcomes

In recent years increased emphasis has been placed on showing the impact of aid expenditure through "rigorous" approaches like randomized controlled trials (RCTs) based on counterfactual logic. One of the claims is that these approaches can assess and attribute the impacts of specific programs separate from other contributory factors. But as Riddel (2012) has pointed out, RCTs are of little value when it comes to evaluating sustainability. They have been successful in evaluating the impacts of specific inputs or pilot projects, but even when they showed positive impacts, RCTs could not assess the prospects of continuing or scaling up an intervention. They can only answer contingent, setting-specific causal questions, such as, "did it work there and then" and cannot be used for generalization to other settings and timeframes unless accompanied by more fine-grained knowledge on the causal mechanisms.

Stern et al (2012) have equally addressed the limitations of counterfactuals and propose a broader framework for designing an evaluation. Evaluation questions should be at the core of the process and the choice of evaluation design should be aligned with the questions and the attributes of the development program. They posit that evaluations typically ask four causal questions:

1. To what extent can a specific (net) impact be attributed to the intervention?
2. Did the intervention make a difference?
3. How has the intervention made a difference?
4. Will the intervention work elsewhere?

The evaluation of sustainable outcomes will most likely deal with combinations of questions 3 and 4 and the underlying assumptions and suitable designs for such an evaluation can be summarized as follows

TABLE 1: SUMMARY OF MOST LIKELY APPLICABLE QUESTIONS, ASSUMPTIONS AND SUITABLE DESIGNS (BASED ON STERN ET AL 2012)

Question	Underlying Assumptions	Suitable Designs
How has the intervention made a difference?	(1) Interventions interact with other causal factors. (2) It is possible to clearly represent the causal process through which the intervention made a difference	(1) Theory-based evaluation especially "realist" variants (2) Participatory approaches
Will the intervention work elsewhere?	(1) What has worked in one place can work somewhere else. (2) Stakeholders will cooperate in joint donor / beneficiary evaluations.	(1) Participatory approaches (2) Natural experiments (3) Synthesis studies

- **Theory-based evaluations** map out an intervention's theory, i.e., the causal chain from inputs to outcomes and impact, and test the underlying assumptions (Funnell and Rogers, 2011). They take into account the social, political and cultural context. A program theory should be elaborated in a flexible way, ready to adapt to changing circumstances and to take on board competing theories and unintended consequences (White 2009). The realist variant asks 'what works for whom and under which context conditions,' for which it makes use of Context-Mechanism-Outcome configurations (Pawson and Tilley, 1997).

- **Participatory approaches** are based on core development ideas like partnership, ownership and democratic accountability. They incorporate stakeholder perspectives to the conclusions and lessons learned from an evaluation of impact (Feuerstein, 1986), assuming that their involvement will increase ownership and the likelihood of continued success. Participatory approaches to causal inference see recipients of aid as active agents that can help "cause" successful outcomes by their own actions and decisions (Chambers 2008).

- **Natural experiments** mean to include experimental elements in program design, e.g. tests out pilot interventions in different alternatives in clusters of different settings before scaling them up. In such cases, the experiment is about means rather than ends, taking advantage of a diversified set of program designs in order to learn through their implementation. Such a strategy is suited to customized, decentralized initiatives that are attempting to achieve similar goals.

- **Synthesis studies** try to accumulate lessons across a set of interventions. Common forms are meta-analysis or narrative synthesis, the basis for causal interference is accumulation and aggregation of evidence within a number of perspectives.

Before deciding on a particular design, program attributes should be taken into account as well. Situational responsiveness has lately gained prominence in evaluation, i.e. matching an evaluation approach to an intervention's degree of complexity (Patton, 2010; Rogers, 2008). Three types of situations can be distinguished that have different characteristics and causal patterns:

- **Simple:** there is high certainty and high agreement between stakeholders about what to do. Cause and effect patterns are clear, predictable and controllable.

- **Complicated:** there is some disagreement about what to do and some uncertainty (due the involvement of many actors or areas of expertise). The relationships between cause and effect depend on the context, they are neither obvious nor predictable and there are alternative routes to achieve affects.

- **Complex:** there is high uncertainty and high disagreement between stakeholders about what to do. The relationships between cause and effect are only evident in retrospect and depend heavily on initial conditions.

These distinctions provide a heuristic for situational recognition, but it is important to bear in mind that "situation" refers to aspects of an intervention rather than the interventions themselves. Hence a complex intervention may have some simple aspects or simple programs may have isolated complex features – and the design implications might be within an intervention rather than for an entire intervention. Stern et al (2012) have used a different and more refined set of program attributes that takes into account impact trajectories, uncertainty and risk, degree of standardization and targeting. They have outlined the evaluation challenges and design implications for their own set of attributes as well as the contingency categories mentioned above.

Design refers to the overarching logic of how an evaluation of impact is conducted, i.e. forms of theory and uses of data that support causal inference. Each of these designs can be associated with a range of methods and techniques, whereby different designs may share similar methods. Below is an outline of methods appropriate to the above design options, which are in line with a systemic understanding of sustainability and seem to be particularly appropriate for ex-post evaluations of sustainable outcomes.[3]

Appropriate Methods for Ex-Post Evaluations of Sustainable Outcomes

Realist Evaluation: A realist evaluation of impact compares whether a program works differently in different localities (and if so, how and why) or for different population groups (Pawson and Tilley, 1997). Realism is based on a "generative" understanding of causation whereby the interaction between what a program provides and the reasoning of its intended target population causes the outcomes. This interaction constitutes a "mechanism" and whether it operates at a particular moment depends on context. Thus hypotheses about Context-Mechanism-Outcome (CMO) configurations are at the core of this evaluation approach. A realist approach is not a method but a way of thinking, so realist design can be incorporated within almost any evaluation of impact. Realist evaluation (Westhorp, 2014) is particularly appropriate for evaluating:

- New initiatives, pilot programs and trials, or programs that seem to work but 'for whom and how' is not yet understood.
- Programs that will be scaled out, to understand how to adapt the intervention to new contexts.
- Programs that have previously demonstrated mixed patterns of outcomes, to understand how and why the differences occur.

Realist Synthesis: is an approach to review evidence on complex social development interventions that provides an explanatory analysis of how and why they work in particular contexts or settings (Pawson et al 2004). Theoretical thinking and empirical evidence are collected and assessed in a heterogeneous and iterative process. In the beginning the theoretical underpinnings of an intervention and the rationale of the review are identified and this framework is then used for locating, integrating, comparing and contrasting empirical evidence. The work is documented in a transparent manner and commissioners or decision-makers are closely involved in shaping the conclusions and recommendations to be drawn from the review. This form of synthesis is particularly suited in cases where an intervention has been carried out in various sites, the theory is not well defined and multiple sources of data or information are available as evidence.

Contribution Analysis has been developed for situations of multiple causation, where several causes can lead to an effect and attribution is either not possible or not feasible (Mayne 2008). Capturing and validating the contribution of an intervention to intended effects and the progress in achieving them infer causality. This enables establishing that an intervention is a necessary part of a strategy, which is sufficient to achieve an intended change, but in combination with other factors. Data on the achievement of effects is verified against the intervention theory, validated by different stakeholders and synthesized in the form of a contribution story. To establish the plausibility of these contributions alternative explanations of effects are investigated. If rival explanations cannot be excluded, further evidence for the contributions is sought. Contribution Analysis requires a sufficient number of cases with comparable characteristics and can identify causation typologies. It is increasingly being implemented in aid interventions and is gaining recognition as a viable alternative to counterfactual evaluations (Journal Evaluation 2012).

[3] Since these designs and methods cannot be directly related to definitions of sustainable outcomes previously described (which only provides a range of possible definitions) it is neither feasible nor useful to specify such links.

Two methods recommended for Complexity Aware Monitoring (USAID, 2013) can also be used for evaluation purposes, especially for complicated and complex aspects of interventions.

Outcome Harvesting: This participatory method inspired by Outcome Mapping understands outcomes as behavioral changes, which makes it particularly suited for assessing capacity building efforts. Outcome harvesting emphasizes utilization-focused practice and enables users to identify, verify and make sense of outcomes with or without reference to predetermined objectives. When used in this latter sense it is appropriate for identifying unplanned outcomes. On the basis of actionable questions (defined jointly with primary users) information is collected and synthesized in the form of outcome descriptions, which explain changes that have occurred in social actors and how change agents contributed to these changes. Evidence of outcomes is validated or substantiated with independent sources, and for each change plausible cause-effect explanations are established for an intervention's contribution. Finally, outcomes are analyzed and interpreted to answer the actionable questions, by considering three perspectives - primary users, change agents and substantiators (Wilson-Grau et al, 2013).

Process Monitoring of Impacts: This method identifies processes considered relevant for the achievement of effects and then verifies whether these processes are valid and actually take place. The processes are located between the results in a LogFrame or results framework and describe how a result at one level is used (by specific actors) to achieve results at the next level. First a logic model is drawn up that connects the outputs to the various levels of intended effects, and assumptions can be inserted about the use of outputs. The logic model may also include context factors that affect the achievement of effects, as well as their relationship with elements of the intervention (e.g. results). Taken together, they bound the "area of observation" that is considered most critical to project success, and the processes and interrelationships within this area are then verified in an evaluation. Thus the likeliness of future outcomes can be assessed by validating the preceding step. The method is particularly appropriate in cases where predominantly output data is available, and can provide proxy estimates on sustainable outcomes by investigating the use of these outputs as a necessary, but not sufficient step for achieving them (Williams and Hummelbrunner 2011).

CONCLUSIONS

This literature review demonstrates that in the field of social development there is a diversity of opinion with respect to the understanding of sustainability and how best to evaluate it. There are different (and sometimes incompatible) ways of understanding the key terms "sustainability" and "sustainable outcomes" as well as the processes contributing to it. However, there are also some patterns that emerge from the literature, which might help in deciding on appropriate definitions and evaluative approaches.

- There is an increasing interest in a dynamic understanding of sustainability, which implies adaptive and developmental approaches as well as the importance of system-wide change processes. This is confirmed by lessons learned from recent evaluations of USAID support to basic education.

- There is a recurrent emphasis on the dynamics and drivers of sustainability, i.e. the major factors that have an influence as well as the processes that precede it, which respectively lead to sustainability. In addition, some attempts have been made – notably in health and social development – to conceptualize these dynamics in the form of frameworks.

- There is a tendency towards planned approaches to sustainability and pathway models for defining sustainable outcomes, which foresee various possibilities to achieve them. These provide various options addressing sustainability in evaluations (e.g. at the level of activity, capacity, ideas).

- There is a growing interest in a systemic understanding of sustainability, which emphasize the interaction between an intervention and its context (including wider policy systems). Some of the most recent sustainability frameworks are explicitly based on systems ideas.

With respect to the evaluation of sustainable outcomes there are some (alas not many) lessons that can be built upon, notably with respect to the role and focus of evaluation as well as measurement and data collection. A framework is recommended to guide the development of evaluation designs and the choice of appropriate methods. Suitable design options are outlined and several evaluation methods identified that are in line with a dynamic as well as systemic understanding of sustainability. These methods have proved to be credible and feasible alternatives to counterfactual evaluations, which are considered of rather limited value for this evaluation task.

Finally, several contributions from the systems field have been identified that can be used in the forthcoming evaluation series. Some are located at the conceptual level (i.e. non-triviality, viability, feedback, complex adaptive systems), others at the level of methods and techniques: several visualization tools were outlined that are widely known and many evaluators are familiar with (feedback loops, causal loop diagrams, stock-flow-diagrams). Other methods – notably those associated with the systems concepts "perspectives" and "boundaries" – would also be available (e.g. for identifying alternative framings or dealing with different perspectives) but would demand specialized skills and expertise.

Implications for Ex-Post Evaluation Series

The task ahead now is to reflect on the relevance of this rather broad menu for the specific aims and needs of the forthcoming evaluation series focusing on sustainable outcomes in basic education, which is to be carried out in a partnership arrangement between several teams/working groups within and outside USAID. Since there are obviously no off-the-shelf methods available for this evaluation series, this reflection and the corresponding choices should be guided by creativity, aiming for mixes of methods or techniques that are appropriate for the specific situation (and thus might differ for project types or the time elapsed after an intervention). The fact that this evaluation series can make comparisons across programs and contexts provides a good opportunity for establishing a coherent and systematic knowledge base about program sustainability.

ANNEX A: BIBLIOGRAPHY

Axelrod, Robert and Cohen, Michael D. (2001). Harnessing Complexity: Organizational Implications of a Scientific Frontier. Basic Books.

Britt, H. (2013). Complexity-Aware Monitoring Discussion Note. Washington, DC: U. S. Agency for International Development.

Chambers, Robert; Karlan, Dean; Ravallion, Martin and Rogers, Patricia (2009). Designing impact evaluations: different perspectives. International Initiative for Impact Evaluation (3ie), Working Paper 3. New Delhi.

Chapman, David and Quijada, Jessica Jester (2009). "An analysis of USAID assistance to basic education in the developing world, 1990–2005." International Journal of Education Development.

Chianca, Thomaz (2008). The OECD/DAC Criteria for International Development Evaluations: An Assessment and Ideas for Improvement. Journal of MultiDisciplinary Evaluation, Volume 5, Number 941

European Court of Auditors (2011). EU Development Assistance for Basic Education in Sub-Saharan Africa and South Asia, Special Report No. 12/2010, European Union, Luxembourg.

Foerster, Heinz von (1961). A Predictive Model for Self-Organizing Systems, Part I: Cybernetica 3, pp. 258–300; Part II: Cybernetica 4, pp. 20–55.

Forrester, Jay W. (1961). Industrial dynamics. Waltham, MA: Pegasus Communications.

Forrester, Jay W. (1993). "Systems Dynamics and the Lessons of 35 years". In Kenyon B., De Greene (ed.). Systems-based Approach to Policy making. Kluwer Academic Publishers.

Feuerstein, Marie-Therese (1986). Partners in Evaluation: Evaluating Development and Community Programmes with Participants. Macmillan Publishers Ltd.

Funnell, S. And Rogers, P. (2011) Purposeful Program Theory: Effective Use of Theories and Change and Logic Models. San Francisco: Jossey-Bass/Wiley.

Gillies, J. (2010). The Power of Persistence: Education System Reform and Aid Effectiveness, Case Studies in Long-Term Education Reform, USAID, EQUIP2.

Glewwe, P., and M. Kremer (2005). 'Schools, Teachers, and Education Outcomes in Developing Countries, CID Working Paper No. 122, Center for International Development at Harvard University, Cambridge, MA.

Gunderson, L. and Holling, C.S. (Eds.) (2002). Panarchy: Understanding transformations in human and natural systems. Island Press. Washington DC.

Hummelbrunner, Richard. 2007. Systemic Evaluation in the Field of Regional Development. In Systems concepts in evaluation: An expert anthology, ed. Bob Williams and Iraj Imam. Point Reyes: American Evaluation Association/EdgePress.

Journal Evaluation (2012). Special issue on Contribution Analysis. Volume 18, Number 3.

Levine, Simon (2014). Assessing resilience: why quantification misses the point. Overseas Development Institute. London. HPG Working Paper

Mancini, J. and Mrek L. (2004) 'Sustaining Community-Based Programs for Families: Conceptualization and Measurement', Family Relations (53): 339-347.

Mayne, John (2008). Contribution analysis: An approach to exploring cause and effect Institutional Learning and Change (ILAC) Initiative. ILAC Brief.

McKeown, Max (2012) Adaptability: The Art of Winning, Kogan Page.

Meadows, Dennis and Donella H. L., Randers ,Jorgen, Behrens ,William W. III (1972). The Limits to Growth. Universe Books. New York.

Meadows, Donella and Wright, Deanna (Ed.) (2008). Thinking in Systems: A Primer. Chelsea Green Publishing.

Midgley, Gerald (2000). Systemic intervention: philosophy, methodology, and practice. New York. Kluwer/Plenum.

Morell, Jonathan A. (2010). Evaluation in the Face of Uncertainty. New York: Guilford Press.

Nkansa, G.A. & Chapman, D.W. (2006). Sustaining community participation: What persists after the money ends? International Review of Education, 52(6).

O'Connell, Deborah et al. (2013). Designing for Action: Principles of Effective Sustainability Measurement, World Economic Forum's Global Agenda Council on Measuring Sustainability.

Organization for Economic Cooperation and Development (1992). Development Assistance Manual: DAC Principles for Effective Aid. Paris.

Paina. Ligia and Peters, David H (2011). Understanding pathways for scaling up health services through the lens of complex adaptive systems. Accessed at: http://heapol.oxfordjournals.org/

Patton, Michael Q. 2010. Developmental evaluation. New York: Guilford Press.

Pawson, Ray, and Tilley, Nick (1997) Realistic Evaluation. Thousand Oaks, Calif.: Sage.

Pawson Ray; Greenhalgh, Trisha; Harvey, Gill and Walshe, Kieran (2004). Realist synthesis: an introduction. ESRC Research Methods Programme. Methods Paper 2/2004. University of Manchester.

Pluye P, Potvin L, Denis JL. (2004). Making public health programs last: conceptualizing sustainability. Evaluation and Program Planning ;27(2):121–133

Ramalingam, Ben (2013). Aid on the Edge of Chaos. Oxford University Press. Oxford, New York.

Riddell, Abby (2012). The Effectiveness of Foreign Aid to Education: What Can Be Learned? Working Paper No. 2012/75, United Nations University.

Richardson G. and Pugh A. (1981) Introduction to Systems Dynamic Modeling. Waltham MA: Pegasus Communications.

Richardson, G.P. (1986). "Problems with Causal-loop diagrams". System Dynamics Review 2(2): 158-170.

Rogers, Patricia and Bob Williams (2008). "Sustainability of Services for Young Children and Their Families: What Works?". Australian Research Alliance for Children & Youth.

Rogers, Patricia (2008). "Using Programme Theory for Complicated and Complex Programmes." Evaluation: the international journal of theory, research and practice, 2008, 14 (1), 29–48, Sage. London.

Ruffer, Tim and Elise Wach (2013). "Review of Making Markets Work for the Poor (M4P) Evaluation Methods and Approaches." DFID Working Paper 41. London: Department for International Development.

Sarriot, Eric, Jim Ricca, Jennifer Yourkavitch, Leo Ryan, and the Sustained Health Outcomes (SHOUT) Group (2008). "Taking the Long View: A Practical Guide to Sustainability Planning and Measurement in Community-Oriented Health Programming." Calverton MD: Macro International Inc.

Sarriot, Eric and Kouletio, Michelle (2014). Community Health Systems as Complex Adaptive Systems: Ontology and Praxis Lessons from an Urban Health Experience with Demonstrated Sustainability. Systems Practice Action Research. Accessed at: http://link.springer.com/

Sarriot, Eric (2104). Advancing the application of systems thinking in health: sustainability evaluation as learning and sense-making in a complex urban health system in Northern Bangladesh. Health Research Policy and Systems 2014, 12:45.

Savigny, Don de and Adam, Taghreed (Eds.) (2009). Systems thinking for health systems strengthening. Alliance for Health Policy and Systems Research. World Health Organization.

Scheirer, Marie Ann (2005) 'Is Sustainability Possible? A Review and Commentary on Empirical Studies of Program Sustainability' American Journal of Evaluation, Vol. 26 No. 3, September 2005 320-347.

Scheirer, Marie Ann and Dearing, James W. (2011). An Agenda for Research on the Sustainability of Public Health Programs. Am Journal for Public Health. November; 101(11): 2059–2067.

Schroeter, Daniela (2010). Sustainability Evaluation Checklist. PhD dissertation, The Evaluation Center, Western Michigan University. Accessed at: http://www.wmich.edu/evalctr/wp-content/uploads/2010/06/SEC-revised1.pdf

Shediac-Rizkallah, M. C., & Bone, L. R. (1998) Planning for the sustainability of community-based health programs: Conceptual frameworks and future directions for research, practice and policy. Health Education Research, 13, 87- 108

Snowden, David (2012). Exaptation & managed serendipity at: http://cognitive-edge.com/blog/entry/5573/exaptation-managed-serendipity-part-i/

Sterman, J.D. (2000). Business Dynamics: Systems Thinking and Modeling for a complex world. Boston MA. Irwin/McGraw-Hill

Stockmann, Reinhard (1993). Hilft die Entwicklungshilfe langfristig? Bestandsaufnahme zur Nachhaltigkeit von Entwicklungsprojekten. Westdeutscher Verlag. Opladen.

Westley, Frances and Miller, P.S. (Eds.) (2003) Experiments in consilience: Integrating social and scientific responses to safe endangered species. Island Press. Washington DC.

Weiss, H., J. Coffman, and Marielle Bohan-Barker (2002). Evaluation's Role in Supporting Initiative Sustainability. Harvard Family Research Project.

Westhorp, Gill (2014). Realist Impact Evaluation: An Introduciton. Overseas Development Institute (ODI), Method Lab Series. London.

Williams, Bob, and Iraj Imam (Eds.) (2007). Systems concepts in evaluation: An expert anthology. Point Reyes: American Evaluation Association/EdgePress.

Williams, Bob and Richard Hummelbrunner (2011). Systems Concepts in Action : a practitioners toolkit. Stanford, CA: Stanford University Press

Wilson-Grau, Ricardo; Rassmann, Kornelia; Smith, Richard and Mauremootoo, John (2013). Retrospective 'Outcome harvesting'. Better Evaluation. Accessed at: http://betterevaluation.org/sites/default/files/Retrospective%20outcome%20harvesting.pdf

White, Howard (2009). Theory-Based Impact Evaluation: Principles and Practice. International Initiative for Impact Evaluation (3ie), Working Paper 3. New Delhi.

ANNEX B: WEBSITES

http://betterevaluation.org/blog/evaluating_sustainability

BetterEvaluation is an international collaboration to improve evaluation practice and theory by sharing and generating information about options and approaches. Contains several blogs on evaluating sustainability and a range of material.

http://www.3ieimpact.org/

The website of the International Initiative for Impact Evaluation (3ie) is an international platform on impact evaluation in development. It offers a range of publications, a newsletter, several blogs, and information on evidence, funding, or training opportunities for impact evaluation.

http://www.cedarscenter.com/

The website of the Center for Design and Research in Sustainability (CEDARS) has an extensive Resource Center on sustainability in Health and Human Development, material on the Sustainability Framework, and a blog on sustainable health and social development.

http://learningforsustainability.net/willallen.php

This website from Will Allen provides a range of resources that focus on monitoring and evaluation that leads to sustainable development. The resources include guides, frameworks and tools for change along with research methods and approaches.

http://www.sustainingability.org

This website presents technical information on the sustainability of health and social systems as well as methods for analyzing, implementing and monitoring sustainability complete with case studies.

ANNEX C: DOCUMENT SEARCH TERMS

The following search terms were used by the reviewer in conducting searches for materials used in this review:

- Sustainable Outcomes
- Viability Of Effects
- Sustainable Results
- Sustaining Learning
- Sustainability In Development
- Sustaining Effects Of Development Aid

ANNEX D: STATEMENT OF WORK

Statement of Work (SOW)
Evaluating Sustainable Outcomes Literature Review for the Sustainable Outcomes Ex-Post Evaluation Series on Basic Education

1. Task Description

This Statement of Work (SOW) outlines the requirements for a review of literature addressing the evidence or evidence gaps on the factors that contribute to the sustainability of outcomes in basic education programming. This literature review will inform a series of ex-post evaluations on sustainable outcomes in the basic education sector currently being designed.

2. Background

USAID's Office of Learning, Evaluation, and Research, in the Bureau for Policy, Planning, and Learning (PPL/LER), is undertaking a series of ex-post evaluations of USAID programs concerning "sustainable outcomes." This purpose of the evaluation series is to better understand what USAID and its implementing partners can do to promote sustainable outcomes. The series will take a systemic approach to understand a variety of perspectives and intended and unanticipated outcomes. Basic education activities will be the evaluand for an initial round of evaluations under the sustainable outcomes rubric, and provide the context for site selection and determining which activities and outcomes to focus on as a starting point. The E3 Analytics and Evaluation Project, led by Management Systems International (MSI) along with team members Development & Training Services (dTS) and NORC at the University of Chicago, is supporting the design and implementation of this evaluation series.

This evaluation is taking place in the context of a paradigm shift in the donor community that is placing greater emphasis on the sustainability of development outcomes in all contexts, but particularly in fast-moving and dynamic contexts such as those found in conflict-affected and post-conflict areas. This shift is represented by the White House's Fact Sheet "Fact Sheet: U.S. Global Development Policy" (September 22, 2010) and USAID Policy Paper "Local Systems: A Framework for Supporting Sustained Development" (April 2014).

3. Objectives of the Literature Review

The objectives of the literature review are to identify existing research, literature and evidence on the use of the terms "sustainability" and "sustainable outcomes" (Objective 1), and lessons that have been learned about evaluating sustainable outcomes (Objective 2).

Objective 1

The first objective of the literature review will be to address the questions:

- What is sustainability?
- What is the relationship between sustainability and adaptability?
- What does the systems field contribute to this discourse?

- How might "sustainable outcomes" be defined?

Objective 2

The second objective of the literature review will be to address the questions:

- What lessons have been learned about evaluating sustainable outcomes that might apply to this evaluation series – sustainable outcomes in basic education?

Subsidiary issues that should be addressed relating to this main topic are:

- What methods are appropriate for ex-post (post-completion) evaluations of sustainable outcomes?
- What are the links between evaluation approaches and the definitions of sustainable outcomes addressed though Objective 1?

The literature review will, on the one hand, summarize current thinking in the systems field with respect to sustainable outcomes, including definition(s) of the term, dimensions of sustainability as well as key assumptions and factors underpinning the pathways to sustainable outcomes. On the other hand, an overview will be given of appropriate evaluation approaches for sustainable outcomes, highlighting evaluation questions and the sustainability dimensions they are best suited to address.

All of this will be substantiated by referencing the respective literature, which can also be consulted for designing and conducting the evaluation series.

4. Data Sources

The review will draw on a range of documents, including academic articles, working papers, discussion papers, book chapters and books, case studies and relevant grey literature.

The reviewer should document all search criteria/terms and databases used to identify relevant literature so that the review can be duplicated.

This literature review will be based upon a curated selection of documents that will be most helpful for designing this evaluation within the boundaries of sustainability and systems. This review is not intended to be based on a comprehensive search of all materials relating to the topics.

5. Audience, and Intended Use

The primary audience for this task will be PPL/LER staff involved in the design of this evaluation series, members (to be identified) of a Methods Working Group, the E3 Analytics and Evaluation Project team supporting the evaluation design process, and the eventual evaluation team(s) that conduct the evaluation series.

The literature review will inform future decisions about the scope and parameters of the ex-post evaluation series. It is also anticipated that the literature review will serve as an ongoing resource for the research process over the course of the evaluation series. It could be used multiple times, adapting to different analytical routes.

6. Deliverables

The following deliverables are anticipated under this task:

Deliverable	Estimated Due Date
Draft Literature Review	o/a December 5, 2014
Final Literature Review	o/a two weeks following receipt of USAID comments on the Draft Literature Review

The Literature Review should include the following sections:

- Executive Summary
- Methodology
- Findings
- Conclusions
- Bibliography

The Final Literature Review should not exceed 20 pages (excluding footnotes).

Unless an exception is made by USAID in writing, all documents and deliverables under this SOW are expected to be kept internal to USAID and the Project team working on the Sustainable Outcomes series.

7. Team Composition

The literature review is expected to be primarily conducted by a subject matter expert, with additional research support as required. Additional home office support, oversight, and expertise will provided by the E3 Analytics and Evaluation Project team.

Subject Matter Expert

The subject matter expert will have extensive experience in: (a) developing, implementing or evaluating international development assistance programs in a variety of contexts; and (b) conducting research or assessments on the sustainability of international development assistance programs. Specifically, the expert will have contributed to the body of knowledge on strategies to promote sustainable outcomes in these contexts and should be able to demonstrate a broad knowledge base and familiarity with concepts in systems thinking. The expert should hold at least a master's degree with at least 10 years of experience in the area of international development assistance in developing countries.