A Theory of Systems Evaluation

Derek Cabrera [1,2,3*]
William M.K. Trochim [1,4,5]

May 25, 2006

Paper appears in:
Cabrera, D. (Ed.). (2006). *Systems evaluation and evaluation systems whitepaper series.* Ithaca, NY: Cornell University Dspace Open Access Repository. National Science Foundation Systems Evaluation Grant No. EREC-0535492.

[*] *Corresponding author. Email address: dac66@cornell.edu*
[1] Cornell University
[2] National Science Foundation IGERT Fellow in Nonlinear Systems
[3] Post Doctoral Associate, Human Ecology
[4] Professor, Policy Analysis & Management
[5] Director of Evaluation for Extension and Outreach

1 Introduction

Systems thinking (in addition to systems methods, systems science, and systems theories that are collectively called systems approaches) is an emerging "field of fields" that is gaining popularity in several arenas (e.g., public health, education, management and organizations, earth sciences, engineering, biology and ecology, complexity, sustainability, and science in general). Systems thinking is a conceptual framework in which a concept about a phenomena evolves by recursively applying rules to each construct and thus changes or eliminates existing constructs or creates new ones until an internally consistent conclusion is reached (Cabrera, 2006). The rules include: 1) *Distinction making*, which is differentiating between a concepts identity (what it is) and the other (what it is not); between what is internal and what is external to the boundaries of the concept or system of concepts; 2) *InterRelating*, which is inter-linking one concept to another by identifying reciprocal (i.e., 2 x 2) causes and effects; 3) *Organizing*, which is lumping or splitting concepts into larger wholes or smaller parts; and, 4) *Perspective taking*, which is reorienting a system of concepts by deciding the focal point from which observation occurs by attributing to a point in the system a view of the other objects in the system (e.g., a point of view).

In evaluation the popularity of systems approaches is growing. Testimony to this trend is the EVAL-SYS listserv, the AEA Systems in Evaluation TIG and the TIG's various sessions at the 2005 AEA Conference. Yet, while there is a great deal of interest in and support for systems approaches in evaluation, there is also much confusion and ambiguity as to what "systems evaluation" means (Cabrera 2006). One goal of the recent NSF EREC grant in building systems evaluation capacity is to develop both a theory of systems evaluation and a protocol that helps practitioners to implement systems evaluation.

The Encyclopedia of Evaluation (Mathison 2005) represents over 600 concepts in evaluation from over 100 contributors. The concepts described in each of these entries may be significantly transformed if viewed from a systems perspective. Of course, many of these entries and contributors already take a systems perspective. Yet, no theory of systems evaluation exists. Development of a theory of systems evaluation (TSE) is an important next step and contribution to the evaluation communities. First, a TSE formalizes both the "systems" concept and the transformative effects it may have on evaluation concepts. Second, a TSE provides a framework for debate and critique that will in turn modify and advance the theory. Third, a vetted TSE provides a framework upon which protocols and practice can be based, spawning new developments in the field. Fourth, a vetted theory, as in other fields such as biology (evolution) and physics (gravity and relativity), provides a general framework within which diverse applications and research can be construed. That is, a vetted general theory allows for both convergence or organization and divergence or diversity.

Some of the more popular evaluation models include theory of change, logic models, causal mapping, and pathway models. Each of these will be discussed in the following section.

2 Popular Evaluation Methods

2.1 Theory of Change

Theory of change (TOC) models use a basic five step process and "backwards mapping" (Figure 1). ActKnowledge, a theory of change company explains, "TOC maps out your initiative through 5 stages:

1. Identifying long-term goals and the assumptions behind them.

2. Backwards mapping and connect the preconditions or requirements necessary to achieve that goal.

3. Identifying the interventions that your initiative will perform to create your desired change.

4. Developing indicators to measure your outcomes to assess the performance of your initiative.

5. Writing a narrative to explain the logic of your initiative," (ActKnowledge & Aspen Institute Roundtable on Community Change 2006).

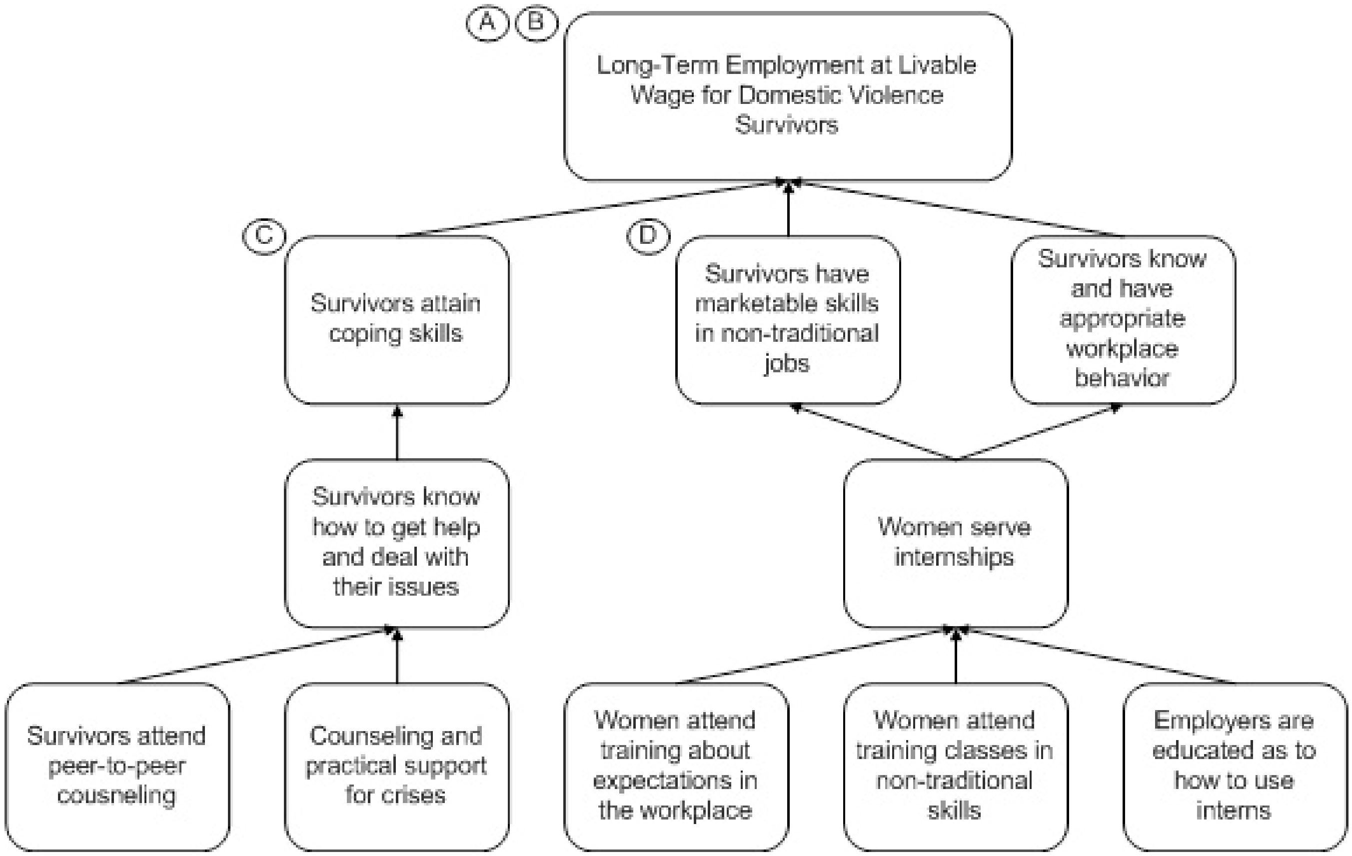


Figure 1: *Theory of change models use "backwards mapping" to "begin with the end in mind," (ActKnowledge & Aspen Institute Roundtable on Community Change 2006)*

TOC models are useful because they force the evaluand to start from large-scale goals and work backwards to interventions that will achieve the goals. In addition, they work with visual "network" maps that link goals to interventions using arrows and "nodes". In addition, TOC models associate "indicators" to these goal-intervention networks. Finally, a narrative is written to explain the "logic" of the initiative. TOC represents, in some ways, a useful tool on the path to a logic model. That is, by visually mapping backwards, TOC eventually brings the evaluand to the point where a logic model can be developed. One might think of the "narrative" as an analog to the matrix of a logic model.

2.2 Logic Models

Logic models use a "spreadsheet" style approach in which a "causal chain" links the situation and environment, inputs, outputs (activities and particpants), and outcomes (short term, mid-term, and long term). Figure 2 illustrates a basic logic model.

2.3 Causal Mapping

Causal mapping uses "directed node and link graphs – similar to concept maps in some ways – to represent a set of causal relationships within a system," (U.C. Berkeley Center for Innovative Learning Technologies

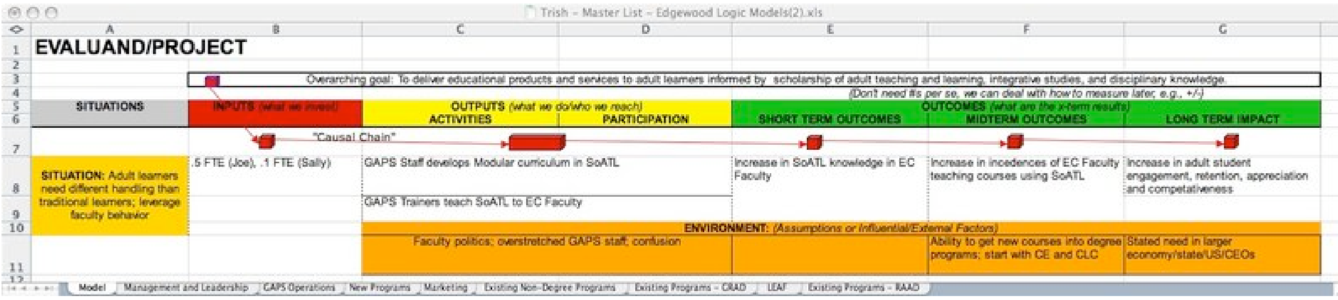


Figure 2: *A "causal chain" links the situation and environment, inputs, outputs (activities and participants), and outcomes (short term, mid-term, and long term)*

2001). Like Novakian concept maps (Concept map 2006), causal maps provide a visual picture of the causal links between one object and another. Unlike Novakian maps however, causal mapping is based on network theory and the mathematical properties of networks. That is, causal links are often associated with a quantity (e.g., a number or a quantitative scale such as increase or decrease, etc.). Causal mapping utilizes only the most basic ideas of networks and is currently used (at least in an explicit way) to map physical systems. Figure 3 illustrates a basic causal map. One can see from Figure 3 that some quantitative analysis can be completed. For example, there are "12 factors and 13 relations" and each relationship can be associated with a number or scale.

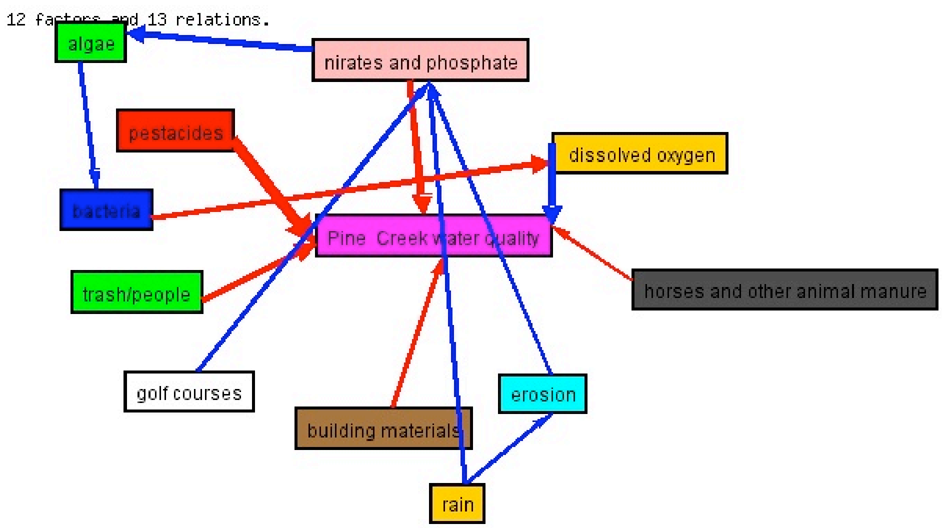


Figure 3: *X increases Y (U.C. Berkeley Center for Innovative Learning Technologies 2001)*

2.4 Pathway Models

Another interesting model is used by the Pathways Initiative. Figure 4 illustrates the basic pathways model. The PMI website states, "PMI does not promote a single formula for family economic success in all communities. We have found no silver bullet. Rather, this website allows users to identify the many arenas of action that contribute to family economic success and to work on the particular elements of the Pathway that are weak or missing in their communities" (Pathways Initiative 2004). Like the other models described here, pathway models link what PMI calls "attributes of effectiveness" (i.e., inputs in logic models) to "actions" that are then linked to sub-goals and goals and finally to outcomes. To a large degree, sub-goals and goals are similar to short-term and mid-term outcomes. The rationale (i.e., "logic") of the program is derived from the actions and sub-goals because these are the actual components of the program. In this way, many action/sub-goal pairings can be linked to the same outcomes; that is, there are many pathways to the same outcome. For communities that want to increase for example, "Family Economic Success," one can imagine numerous program pathways that contribute to this larger goal.

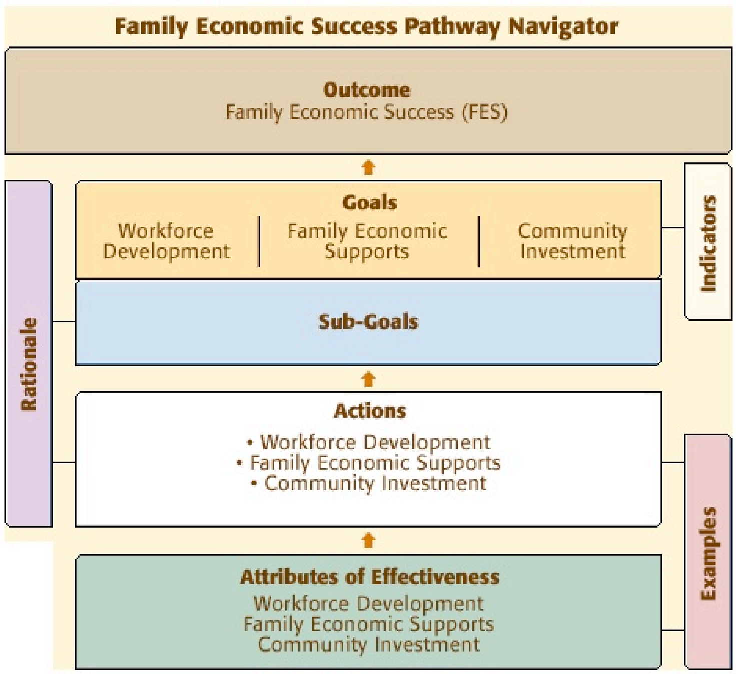
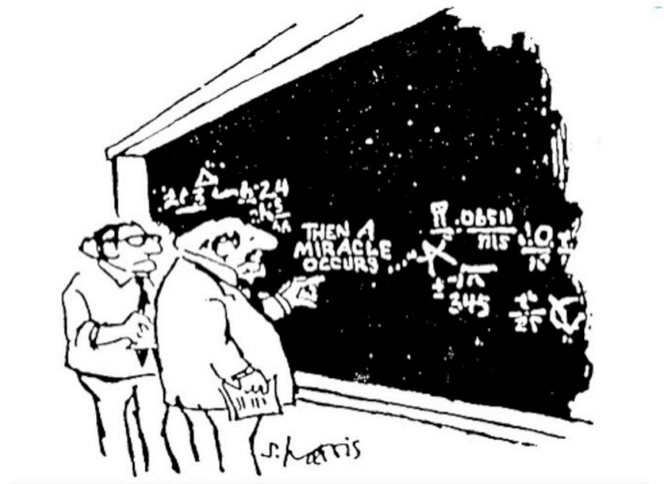


Figure 4: *A Pathway is one "path" that a collective takes to get to an outcome (Pathways Initiative 2004).*

2.5 Conclusions

There is a large degree of agreement between these models. Each model contains aspects of the others and each model contributes a new attribute to our understanding. It is clear that "causality" is a central idea–that some intervention, action, object, or activity (X) leads to some outcome (Y). It is also clear that there are often interim steps (i.e., "short or mid term outcomes", "goals", or "subgoals") that occur between X and Y. Also clear among these models, is the desire to create "causal chains" or "through lines" that are rational and logical. There are two important ideas within this notion: 1) that there is a "public logic" to the linkages we make and 2) that there is a **clear** path that links one thing to the next. That is, there are no "leaps" of logic in which the link between an activity and an outcome requires some unspecified miraculous intervention (Figure 5).

Another similarity between these models is their use of visual maps or illustrations; this may be because planning and evaluation are inherently concept-to-practice exercises, and practice involves groups of people. The visual output of these maps and models may facilitate group understanding and agreement. Some models contribute to areas in which other models lack robustness. For example, causal maps begin to quantify the qualitative in terms that may be analyzed mathematically. This contribution to the mix may



"I think you should be more explicit here in Step Two."

Figure 5: *A "leap of logic" (Harris 1992)*

have real potential in analyzing larger scale initaitives computationally. That is, letting a computer sift through the data to identify key indicators, "markers" or the attributes of "pathways." Causal mapping points toward such advances but it does not come close to utilizing existing mathematical, theoretical and computational knowledge. Advances in network theory, for example, might prove very useful in developing more advanced models that can be used both locally and for large-scale initiatives to "sift through the noise and find signal." Some of these models are pointing toward the need to find simpler indicators for the complexity of the system; perhaps similar to what ecologists call "indicator species" (Indicator species 2006). Computational methods may assist us in identifying these leverage points or "markers" in a system.

More generally, these models seem to point to another important idea. Evaluation is complex. If it were not complex, such models would not be required. There are many stakeholders and perspectives competing for attention. There are many different actions occurring at different scales, and their evaluation cycles are different in their timing. There are also many types of programs, some large and complicated, others small and simple and still others that are simple and large or complex and small. Each of these programs are presumably at different stages of operation and therefore may require different evaluational methods. One final point that warrants discussion and future thought is the degree to which most of these models follow a "right-left" "up-down" linear chain. That is, the relational arrows in these models tend toward a single direction (causal maps are the exception). This may be because of the need to keep things simple, but it also may be an area that needs further thought and development. That is, because evaluation-systems are complex, it must be presumed that they exhibit feedback (nonlinear relationships). Here again, because computational technology is well-suited to analysis of large data sets, it may be a useful tool in identifying important places where feedback may be influencing the evaluation-system.

We conclude this discussion by outlining some of the major points. We view these ideas as a list of both where these models agree, as well as unique contributions that each model provides:

1. causality
2. multiple pathways to the same place
3. must provide public logic or rationale
4. continuous causal chain

5. network structure and function
    (a) computational analysis to find signal in noise
    (b) link qualitative ideas and quantitative numbers/scales
    (c) analysis of the structure and function of these networks
6. indicators or markers
7. visual representations
8. complexity
    (a) many perspectives, stakeholders, scales, timing, program types, evaluational needs
9. feedback

3 A Theory-Protocol of Systems Evaluation

It is perhaps an audacious goal to require that a theory and protocol for systems evaluation include each of the items in the list above. Of course, all theories and models are eventually wrong. Yet, proposing a theory or model, and subsequent protocol for systems evaluation, can help to move the field by providing an object for dialogue and debate, comparison and contrast. Thus, a theory of systems evaluation is proposed herein. The theory of systems evaluation (TSE) is stated as follows:

> **Systems evaluation is the process of creating "coupled models" for the evaluand construct(s) and coupling those models with "coupled methods" to develop an "ecology of feedback methods" for the evaluand construct(s).**

The TSE is a combination-matrix, which simple means that each of the components interact to alter each other. This interaction is illustrated in Figure 6. The six "structural components" of the TSE are: 1) Evaluand Construct; 2) Perspectives; 3) Scales; 4) Cycles; 5) Types; and 6) Phases. These components are interrelated and create a dynamic output. Figure 6 illustrates the dynamic properties of each of the model components. The way they are interrelated is like a "conceptual algorithm" in which each component is "run through" a process. In the following section, the sequence of steps is described in greater detail.

3.1 Step One: Conceptualize the Evaluand Construct Prime

This is the initial "conceptualization" of the evaluand prime (Ce'; see Figure 7). That is, the concept of the evaluand before we do anything to it. Step 1 is the initial step that establishes that there is "something" to evaluate in rough terms. During this step, one might propose that the evaluand construct prime is a program, the focus of assessment is on students and the long term goal is world peace. That is, something that approximates what the evaluand is. In Steps 2 and 3, this "primed" construct of the evaluand will be subjected to numerous questions and processes in order to determine what the "actual" construct of the evaluand is.

3.2 Step 2: Run the Evaluand Prime Through the TSE Matrix of Components

The second step in the process is to run the "evaluand construct prime" through the matrix; this is a "boundary critique" process and leads to the "evaluand construct" (Ce). It should be noted that Ce may be a complex of constructs, rather than a single construct. This critique determines the "boundaries" of the evaluand construct(s). Boundary critique is essentially the process of determining what should fall within and without the scope of evaluation by posing questions and formulating answers based on the six

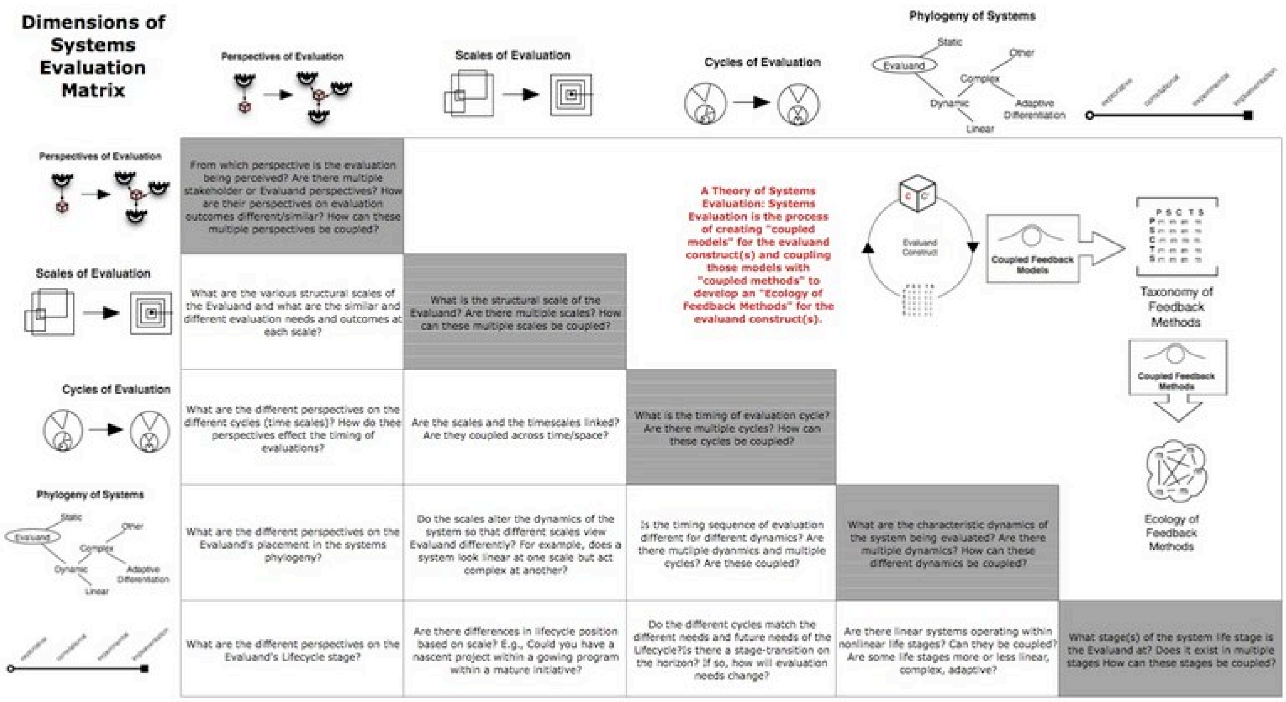


Figure 6: *The dynamic between the components of the theory of systems evaluation*

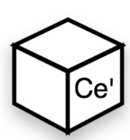


Figure 7: *Ce*

components (perspectives, scale, cycles, types, and phases). The net result of this "boundary critique" is to change Ce' to Ce. Figure 8 shows the transformation of the original construct prime of the evaluand to the final construct of the evaluand. The cycle and table of construct changes (the small c's represent changes to the construct prime) is the process described in Step 2. So, the "little c's matrix" in Figure 8 is a representation of the matrix in Figure 6.

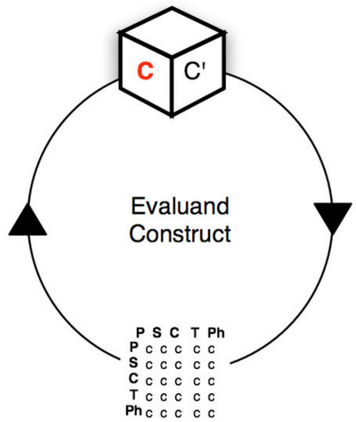


Figure 8: *Boundary Critique*

To perform a boundary critique, one poses questions and formalizes answers based on the *grey boxes* in the matrix (see Figure 6); these boxes represent the individual components. Next, one asks questions relating to the *white boxes* in the matrix; these boxes represent the pairings of components. It may be possible and useful to ask triplets or quadruples of the matrix.

Each component represents an entire conversation about the Ce'. The central theme in each component is transitioning from the component itself to a set of coupled components. So, for example, it is one thing to think in terms of multiple perspectives but it becomes a systemic view when these multiple perspectives are "coupled" so that they logically and rationally work in synthesis. Each component is described in detail below.

Perspectives. **From which perspective is the evaluation being perceived? Are there multiple stakeholder or evaluand perspectives? How are their perspectives on evaluation outcomes different/similar? How can these multiple perspectives be coupled?** Perspectives take multiple forms. For example, individual stakeholders might have different needs, desires or perceived outcomes or activities. Part of the boundary critique is determining which perspectives need to be considered in the evaluand construct and which do not. Note that the other components will help to determine various perspectives. For example, if it is determined that the evaluand is comprised of three hierarchical scales (e.g., NSF-level, program, and project) there may be three unique perspectives associated with these scales (e.g., NSF policy-level administrator, program officer, project director). The image for this component illustrates the need to view the evaluand from not only one perspective but several, and then to ensure that these perspectives are coupled. Coupling across hierarchical scale is central to systems evaluation (see Figure 9).

Scales. **What is the structural scale of the evaluand? Are there multiple scales? How can these multiple scales be coupled?** Like perspectives and cycles, the scale of an evaluand may be multiple and diverse. For example, if you wanted to evaluate a rock climbing program that taught content-knowledge about rock climbing with the goal of increasing self esteem, you might think about different scales. There is the scale of the individual student and aggregate students, the individual instructor and aggregate instructors, comparison between different rock climbing programs at the programmatic scale,

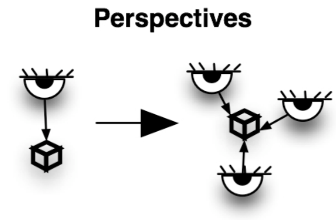


Figure 9: *From a single perspective (left) to multiple, coupled perspectives (right)*

and the comparison between different non-rock climbing programs at the goal-level scale of self-esteem (see Figure 10).

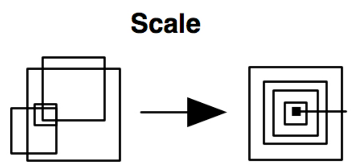


Figure 10: *From many uncoupled hierarchical scales (left) to multiple, coupled scales*

Cycles. **What is the timing of the evaluation cycle? Are there multiple cycles? How can these cycles be coupled?** Feedback and evaluation are synonymous. But not all feedback occurs after the intervention. Organizational learning, for example, is a type of evaluation-feedback that might occur at very small time scales and apply to individuals rather than whole programs. If you wanted to evaluate these micro-scale activities (i.e., learning) the evaluation might occur at a larger timescale. How do these time scales (cycles) couple with each other? There are likely several timescales of evaluation within any given evaluand. Coupling shorter cycles with longer ones is central to systems evaluation (see Figure 11).

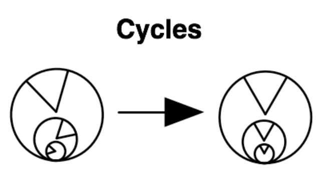


Figure 11: *From multiple, uncoupled evaluation cycles (left) to multiple, coupled cycles*

Types. **What are the characteristic dynamics of the system being evaluated? What are the characteristic dynamics of the types of evaluation that are needed? Are there multiple dynamics? How can these different dynamics be coupled?** Figure 12 suggests a "phylogeny of system types" that can be used to better understand the system and its evaluational needs. For example, can the evaluation be characterized as linear and discrete or is it complex, dynamic and adaptive?

Phases. **What phase(s) of the life cycle is the evaluand/evaluation in? Does it exist in multiple stages? How can these stages be coupled?** For example, is your program a new program exploring new innovative interventions with very little baseline data? If so, an experimental-type

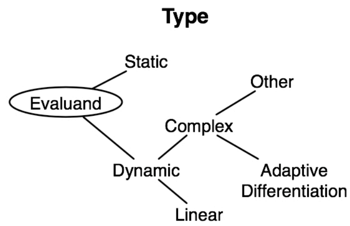


Figure 12: *From one-size-fits-all evaluation to "types-within-types" (e.g., linear within complex, within linear)*

of evaluation may be premature. If your program is more developed and mature, baseline data exists, and a valid construct of your program outcome is available, then an experimental evaluation may be the next step in your program's evolution (see Figure 13).

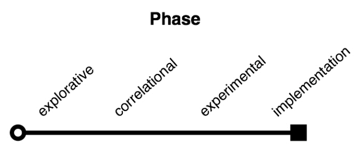


Figure 13: *From evaluation "warfare" to a rationale based on characteristics*

The Output of Steps 1 and 2: An Evaluand Construct and Coupled Feedback Models. Steps 1 and 2 generate an evaluand construct (Ce) that has undergone boundary critique according to the matrix of components. This process leads to the creation of various models for "coupled feedback." For example, one might create a model that couples the feedback from a sub-evaluand that occurs at one time-cycle and a sub-evaluand that occurs at another cycle or at one level of scale and another level of scale. These are called "coupled systems models" because they tell you the various places within your system that need to be linked (see Figure 14).

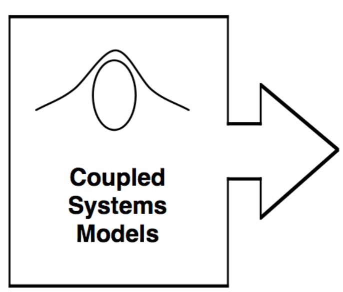


Figure 14: *Coupled mental models of the evaluand*

3.3 Step 3: Converting Coupled Mental Models into Coupled Feedback Models

Note that while you now have a set of coupled models that link multiple perspectives, scales, cycles, types and phases, there are no methodological analogs for these models. That is, while you have now developed a

systemic conceptualization (or mental model) of the evaluation system, you have not created the system itself. The next step is to link these models to evaluation practices and methodologies. The mechanism for doing this step is called **the taxonomy of feedback methods** and it is an evolving categorized taxonomy (categorized using the five components) of many evaluation methods. One can think of this taxonomy as a codified library of methods. Once a coupled mental model of the construct of the evaluand is developed, one needs to link this model to actual feedback methods that can be appropriately used for the types of characteristics of the evaluand. The result of Step 3 is an output called "coupled feedback methods," because the mental model has been linked to methodologies (see Figure 15).

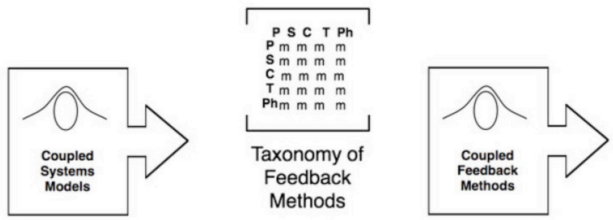


Figure 15: *From mental models of the evaluand to "evaluation Wal Mart" to methodological models of the evaluand*

3.4 Step 4: From a Taxonomy to an Ecology of Feedback

Once the "coupled feedback methods" have been established, the next step is to perform an ecological analysis on the entire set of couplings. This is done qualitatively by developing a linked set of coupled methods of feedback. It is also accomplished quantitatively using ecological evaluation network analysis (EENA).

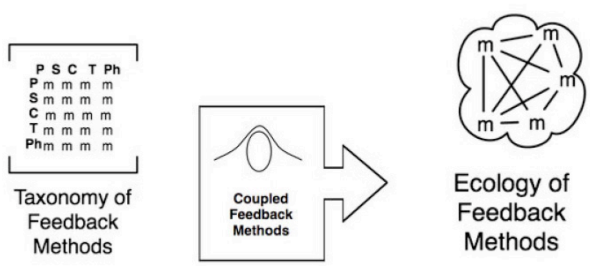


Figure 16: *Toward an ecology of evaluation methods*

Step 4 is both a qualitative and quantitative process that blends the thinking of logic models, causal maps, pathways, and theory of change models into a parent model to each of these models. That is, these models all belong to the set of EENA models. An EENA model has the following structure:

- **A "multigraph" vertice structure**: two types of vertices (activity and outcome). More vertices can be added but most of the network analysis will occur on these two types of vertices. Descriptive statistics could be used for other types of vertices such as inputs, environment, etc.
- **Directed, weighted edges**: the weights are determined by short term outcome (STO=1), mid-term outcome (MTO=2) and long term outcome (LTO=3). In this the incompatible time frames of evaluation are reconciled.

- **Superedge analysis used for pathways: super edges are combinations of tice-edge that can be used to do pathway analysis (not just path, but pathway length, etc.).**
- **"Markers" and "indicators" are defined by the aggregate number of linkages and aggregate edge weight of the vertex. The higher this number, the more of a marker it is for the system**. The most basic EENA is a single activity with a causal chain to three outcomes over time. This is the type of EENA that might be associated with a single activity that is part of a larger project (see Figure 17).

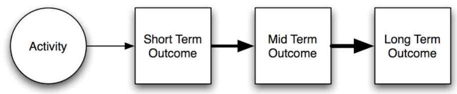


Figure 17: *Basic EENA*

Now imagine that there are numerous activities, each linked in different causal chains but sharing some outcomes at different time scales. Note that different activities share the same outcomes, but also that these outcomes occupy different time periods in the cycle. This means that no outcome is inherently short, mid or long term in an aggregate view when multiple activities are taken into account (see Figure 18).

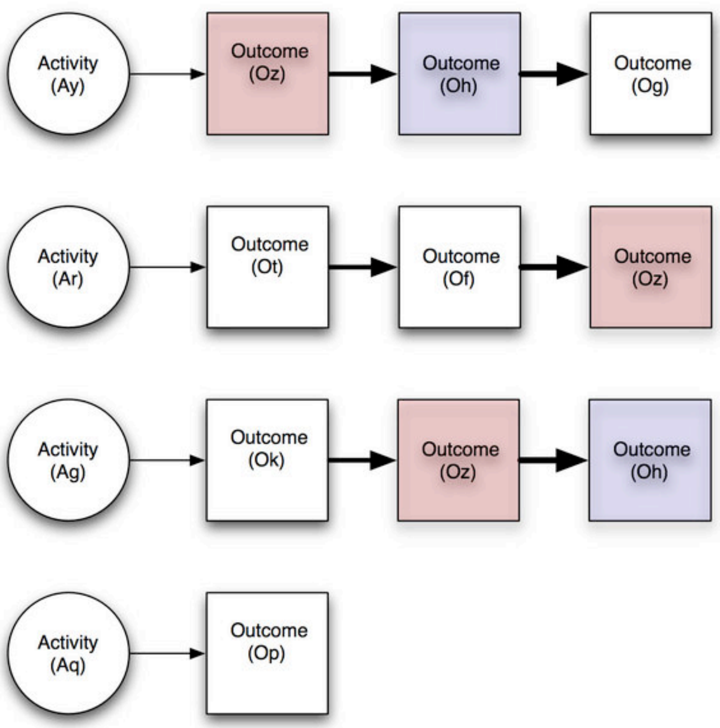


Figure 18: *Multiple activity EENA*

Figure 19 illustrates a more complex but still simple EENA network. In this graph all arrows point to the right, but feedback can be incorporated if necessary. The process of reconciling *same* or *different* outcomes to an ecology of feedback network is not difficult. However, where outcomes are *similar*, the work becomes more difficult. Suffice to say, there are numerous qualitative and quantitative methods (e.g., focus groups, brainstorming, concept mapping, content analysis, or pre-categorization based on baseline data) that can aid evaluators in outcome comparisons and analysis.

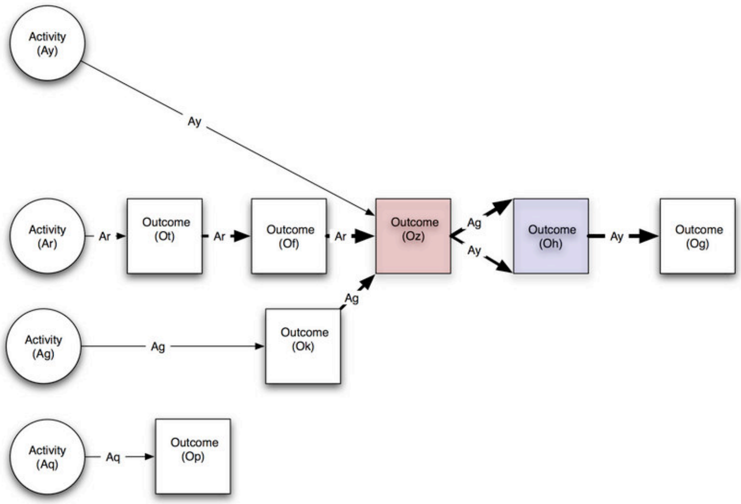


Figure 19: *More complex but still simple EENA network*

Figure 20 describes some of the analysis that can be done using the EENA. There are two types of nodes (activity and outcome) and three line weights; the line weights denote the activity-based outcome ordering. One can see that Oz has different properties from Oh as a marker/indicator. Also, there are numerous pathways, some more direct, to the final outcome Og.

Item	Number	Id
Number of Vertices	11	Ay,Ar,Ag,Aq,Ot,Of,Oz,Oh,Og,Ok,Op
Number of Vertices (vA)	4	Ay,Ar,Ag,Aq
Number of Vertices (vO)	7	Ot,Of,Oz,Oh,Og,Ok,Op
Number of Edges	10	
vA Max Edges	1	Ay,Ar,Ag,Aq
vO Max Edges	3	Oz
vO Max Edge Wt	5	Oz,Oh
vO Max Edge + Max Wt	3/5	Oz
Max Pathway Distance to Og	3	Ay
Pathways to Og	1	Ay

Figure 20: *Analysis using EENA*

4 Bibliography

1. Indicator species. (2006). *Wikipedia, The Free Encyclopedia* Retrieved April 24, 2006, from `http://en.wikipedia.org/wiki/Indicator_species`

2. Concept maps. (2006). *Wikipedia, The Free Encyclopedia* Retrieved May 7, 2006, from `http://en.wikipedia.org/wiki/Concept_map`

3. ActKnowledge, & Aspen Institute Roundtable on Community Change. (2006). Theory of change: Overview. Retrieved April 25, 2006, from `http://www.theoryofchange.org/html/overview.html`

4. Cabrera, D. (2006). *Systems Thinking*. Education. Ithaca, NY, Cornell University. Ph.D. http://hdl.handle.net/1813/2860

5. Harris, S. (1992). *Chalk up another one: The best of sidney harris*. Washington, DC: AAAS Press.

6. Mathison, S. (Ed.). (2005). *Encyclopedia of evaluation*. Thousand Oaks, California; London; New Delhi: Sage Publications.

7. Pathways Initiative. (2004). Pathways to outcomes. Retrieved April 24, 2006, from http://www.pathwaystooutcomes.org/index.cfm

8. U.C. Berkeley Center for Innovative Learning Technologies. (2001). Causal mapping. Cilt synergy project. Retrieved April 24, 2006, from http://cilt.berkeley.edu/synergy/causalmap/