

Adaptive Evaluation

A synergy between complexity theory and evaluation practice

December 31, 2016 · Practitioner

Dr. Glenda Eoyang¹, Judy Oakden

¹ Human Systems Dynamics Institute

Eoyang G, Oakden J. Adaptive Evaluation: A synergy between complexity theory and evaluation practice. *Emergence: Complexity and Organization*. 2016 Dec 31 [last modified: 2017 Feb 25]. Edition 1. doi:

10.emerg/10.17357.e5389f5715a734817dfbeaf25ab335e5.

Abstract

The field of evaluation involves making judgements of quality, value and importance to support accountability assessment, learning, and to improve performance. Traditional evaluation designs assume a high level of predictability and control. The problem is that complex programs or contexts challenge this basic assumption. Often programs deal with emergent outcomes and objectives, adaptive program processes, nonlinear theories of change and evolving stakeholder expectations. Under such complex conditions, traditional evaluation methods and tools do not allow realistic and useful representations of reality. In these instances, we need a more adaptive approach to evaluation. One that fits the environment without compromising rigor. In this paper, we articulate what we have found useful in seeing patterns in complex programs, understanding the dynamics in ways that are meaningful to stakeholders and recommending Adaptive Actions to improve impacts over time. In our work, synergy has emerged between complexity theory (through the lens of human systems dynamics) and evaluation practice (through a case study of a complex program of social change). What emerges at this generative intersection is an evaluation method that is simple, robust, rigorous and flexible enough to meet the demands of twenty-first century social change. We will explore the implications of this approach for theory and practice in complexity and evaluation, and we will share some questions that are emerging for us as we prepare for our next cycle of theory and practice development. In this paper, we provide an overview of the challenge and previous efforts to address it, an introduction to basic theory and practice of human systems dynamics (HSD) and theoretical foundations for a new approach to evaluation in complex environments, Adaptive Evaluation. We demonstrate applications of this new evaluation practice in a case study. Finally, we articulate lessons learned and emerging questions.

Introduction

The field of evaluation involves making judgments of quality, value and importance¹ to support accountability, assessment, learning, and to improve performance. In general, there are three overall distinctions in evaluation: 1) Formative evaluation which is about improving and supporting learning within a programme or initiative, 2) Developmental evaluation which focuses on development and adaption where innovation is occurring and 3) Summative evaluation which focuses on making judgments about the quality, value and importance of the thing being evaluated after the fact².

Some traditional evaluation designs assume there will be high levels of predictability and control. In these cases, program designers and evaluators assume outcomes and objectives can be pre-determined, strategy and processes can be predicted, theories of change are logical and linear and stakeholder expectations remain relatively stable. Based on these assumptions, it is supposed that an evaluation design can be established at the start, and adhering to it is considered part of the rigor of evaluation practice.

The problem is that complex programs or contexts challenge these basic assumptions. Often programs deal with emergent outcomes and objectives, adaptive program processes, nonlinear theories of change and evolving stakeholder expectations. Under such complex conditions, traditional evaluation methods, approaches and tools do not allow realistic and useful representations of reality³. In these instances, we need a more adaptive approach to evaluation. One that fits the environment without compromising rigor.

For the past two decades, a number of professional evaluators^{2,5,10,11}, have explored complexity and systems sciences to develop new ways to see, understand and influence the complex systems they evaluate. As early as 1999, Glenda Eoyang and Tom Berkas³ also speculated about how concepts and tools from complexity science might influence evaluation practice. Since that time, practitioners and academics in the field of evaluation have pushed the bounds of theory and practice. They have moved beyond the constraints of traditional evaluation to provide timely and useful information to their clients about complex, unpredictable systemic change.

Over time, the theory and practice of complex evaluations have evolved. Conflicts among theories and theorists generated and tested new hypotheses. New challenges in practice forced innovation in methods and technologies. Anomalies between theory and practice drove co-evolution of both, as new paradigms of thinking and action emerged. This process is emblematic of a fundamental paradigm shift⁴. The past three decades have witnessed significant change in theories and practices in the fields of

evaluation^{5,2,6} and complexity^{7,8}. This rich dialogue between theory and practice, complexity and evaluation, have created a wide range of options for evaluating complex interactions and emergent outcomes in program implementations.

In this paper, we articulate what we have found useful in seeing patterns in complex programs, understanding the dynamics in ways that are meaningful to stakeholders and recommending Adaptive Actions to improve impacts over time. In our work, synergy has emerged between complexity theory (through the lens of human systems dynamics) and evaluation practice (through a case study of a complex program of social change). What emerges at this generative intersection is an evaluation approach that is simple, robust, rigorous and flexible enough to meet the demands of twenty-first century social change. We will explore the implications of this approach for theory and practice in complexity and evaluation, and we will share some questions that are emerging for us as we prepare for our next cycle of theory and practice development.

In this paper, we provide an overview of the challenge and previous efforts to address it, an introduction to basic theory and practice of human systems dynamics (HSD) and theoretical foundations for a new approach to evaluation in complex environments, *Adaptive Evaluation*. We demonstrate applications of this new evaluation practice in a case study. Finally, we articulate lessons learned and emerging questions.

The problem

Human systems, and the programs designed to improve them, are inherently complex. Patterns of behavior emerge from complex, underlying dynamics of these systems. Those symptomatic patterns, outlined by Eoyang & Berka³, include the complex system being:

1. Dynamic (unpredictable change over time)
2. Massively entangled (nonlinear interdependencies within and across levels)
3. Scale independent (self-similar across scales)
4. Transformative (systemic and persistent change)
5. Emergent (whole is different than the sum of the parts).

These patterns have been expanded by evaluators and others to include the Butterfly Effect (sensitivity to initial conditions), turbulent boundary conditions, unintended consequences and categories of behaviors, including simple, complicated, complex, chaotic. All of these features of complex systems create evaluation challenges, summarized in the table below.

Feature of Complex Systems	Design Challenge	Data Collection Challenge	Analysis/Interpretation Challenge
Dynamic	Unpredictable outcomes/impacts	Shifting indicators and no reliable baseline	Shifting contexts and relationships
Massively entangled	Ambiguous unit of action or unit of analysis	Unpredictable correlation effects	Cultural and contextualized interpretations
Scale independent	Multiple, relevant levels of analysis	Inconsistent units of measure across scales	Lack of focus on one level of action
Transformative	Inconsistent outcomes/impacts	Changes in assumptions	Fundamental questions and contexts change over time
Emergent	Designs are not replicable or transferrable	Unstable time series and shifting units of analysis	Multiple perspectives and interpretations
Butterfly Effects	Distinguishing noise in the system from meaningful signal	Relevance of indicators changes over time	Complex and unpredictable causal relationships
Turbulent boundary conditions	Open systems and multiple-levels of interdependence	Inconsistent methods and measures	Conflicts among groups of stakeholders
Unintended consequences	Incomplete design criteria	Unexpected results	Inability to assign responsibility or causality
Simple, complicated, complex, chaotic	Multiple design alternatives	Diverse methods	Conflicting needs and expectations

Evaluators have long been conscious of the complexity of real programs in real contexts. Models and methods of complexity science have helped them create new models, tools and methods to accommodate the influence of complexity and uncertainty, or to incorporate it into their evaluative thinking.

Current State of Affairs

In the past two decades, there are a number of evaluation practitioners, systems thinkers and complexity theorists who have walked alongside each other learning from each other's work to address the design challenges and to accommodate the complex dynamics that impact on evaluation in many programs and contexts. Each of them focused on one or more of the challenges of complex evaluation, and all of them have proven useful. These diverse perspectives have been integrated into evaluation practice in numerous ways, including:

1. Developmental evaluation²
2. Interrelationships, Boundaries, Perspectives⁵
3. Critical Systems Heuristics⁹
4. Outcome Harvesting¹⁰
5. Evaluative Inquiry¹¹

Patterns not Problems

In our practice, we have found the most effective processes to evaluate change in complex programs and environments will be:

1. Relevant across scales, from individual to community levels of change
 2. Flexible to meet diverse needs in diverse contexts
 3. Adaptable to changes in programs over time
 4. Oriented toward action
-
5. Working with existing data
 6. Useful before, during and after an intervention
 7. Applicable at all the different scales of strategy, policy and practice (precision and accuracy).

In writing this paper, we wondered if framing the dynamics of complex systems, through the lens of human systems dynamics (HSD), combined with an evaluation-specific methodology might create a useful framework that meets these criteria.

The theory

The patterns of complex systems, described by Eoyang & Berkas³, are still true, but the theory and practice of HSD, and its implications for evaluation, have continued to emerge. Today, we understand more about the underlying dynamics that generate these symptomatic patterns, and we have also identified simple principles and practices that influence action to see, understand and influence those patterns.

Dynamics

Complex human systems self-organize. As parts of the system interact, they generate system-wide patterns that affect and are affected by behaviors of individuals. These emergent patterns generate the well-known features of complexity science. If an evaluator understands the underlying dynamics that generate these complex patterns, then they are able to respond with designs, criteria, analysis, synthesis and recommendations that are sensitive to complexity and still easy for stakeholders to understand.

Through the work of HSD, we recognize three basic systemic conditions that define the patterns in complex contexts and influence the speed, path and outcomes of their self-organizing processes¹².

Containers (C) establish the boundaries of a notional system and its subsystem parts. In a complex human environment, there can be many boundaries that are relevant. Some can be explicit, stable and impermeable, while others are implicit, unstable and permeable.

The second condition that makes patterns explicit and also influences them, are the **Differences (D)** within the Containers. Differences, like evaluative criteria, can be few in simple systems and multiple in complicated ones. Over time, differences that are foregrounded in the systemic pattern may disappear to be replaced with others that were previously not visible.

Finally, the parts of complex systems are connected to each other. The pathways along which the parts of the system influence each other are called **Exchanges (E)**. Exchanges can be unidirectional, linear and causal; or they can be mutual, nonlinear and responsive.

These three systemic features, together known as the CDE Model, serve two functions for evaluators. First, they manifest the patterns of performance that are of interest to the program. If an evaluator can see the Containers, Differences and Exchanges in the current state, they can assess whether and how each condition is serving the intended purpose of an intervention. As the parts of the system interact and evolve within and between themselves, it is possible to observe changes in the conditions, even when local and/or systemic transformations cannot be predicted or controlled.

Second, these three conditions determine the pattern, so any shift in one of them will result in the emergence of a new pattern. This new pattern—an outcome of the self-organizing process—can be assessed against the intended or anticipated outcomes. Subsequently, the assessment can inform action to influence the conditions for the future.

In these two ways, patterns (captured as CDE) allow evaluators to make meaning of individual and systemic behaviors before, during and after a strategy, policy or program intervention. This insight, also feeds into practical options for action.

Options for action

Using this CDE Model, the evaluator is able to see current patterns, focus on most relevant conditions, monitor how those conditions change over time and recognize and report on new patterns as they emerge. Based on this understanding of complex dynamics, three practices support engagement with, and evaluate change in, complex systems: Adaptive Action, Simple Rules and Inquiry.

Adaptive Action is a three-step inquiry cycle that allows the evaluator to observe, assess and recommend action as change progresses in the complex system. The process consists of three questions:

1. **What** is the current pattern (container, difference, exchange)?
2. **So what** are the significant tensions in the pattern (given the current containers, differences and exchanges)?
3. **Now what** actions will change the conditions and shift the pattern toward greater quality, value and importance?

This process is also one of the key inquiry approaches used in Developmental Evaluation² and supports its emergent, adaptive processes. In addition to representing the whole evaluation as a learning cycle, Adaptive Action also encourages multiple learning and adaptation cycles for different parts of the system at different timescales and scopes simultaneously.

Simple Rules are guides for local, individual actions that generate system-wide patterns. Based on computer simulation models of group behavior, Simple Rules provide group coherence while supporting maximum freedom for individuals to interpret and implement the rules based on their own local information. Based on our theory and practice, to be effective, a set of Simple Rules should:

1. Include seven or fewer (so they can stay in short-term memory)
2. Begin with an active verb (so they inform observable action)
3. Apply to everyone in the flock (so a pattern emerges across the system)
4. Determine at least one container, one difference and one exchange (to set conditions for a coherent pattern).

A short list of Simple Rules may exist before or emerge in the course of the evaluation. Either way, they can provide insights to inform evaluative activities⁸.

Inquiry is a stance of humility and openness to the unknown. It is imperative for the evaluator of a complex program in emergent settings. Unless evaluators are able to transcend their assumptions and embrace the uncertainty of self-organizing patterns, they will not be able to see, much less evaluate, the quality, value and importance of a social intervention. In HSD, we use four Simple Rules to operationalize this stance of inquiry:

1. Turn judgement into curiosity
2. Turn conflict into shared exploration
3. Turn defensiveness into self-reflection
4. Turn assumptions into questions.

Together, these three HSD practices—Adaptive Action, Simple Rules and Inquiry—provide a useful foundation for an evaluator

to work effectively in complex environments. Blending these practices with an evaluation-specific methodology has inspired the development of a new evaluation approach.

Introduction to an evaluation-specific methodology

Before we move into the evaluation case we provide a brief primer on evaluative logic, an evaluation-specific methodology and define *Adaptive Evaluation*.

What is evaluation logic? Evaluation uses probative inference, a particular kind of logic to make evaluative judgements¹³. Probative inference is where professional judgement, based on evidence, is used to reach a conclusion “beyond legitimate doubt by an objective, reasonable and competent expert”¹³.

What is evaluation-specific methodology? Nunns, Peace, & Witten ¹⁴ explain that to be evaluative, evaluators need to establish criteria, construct standards, measure performance and then come to evaluative conclusions. This is an evaluation-specific methodology.

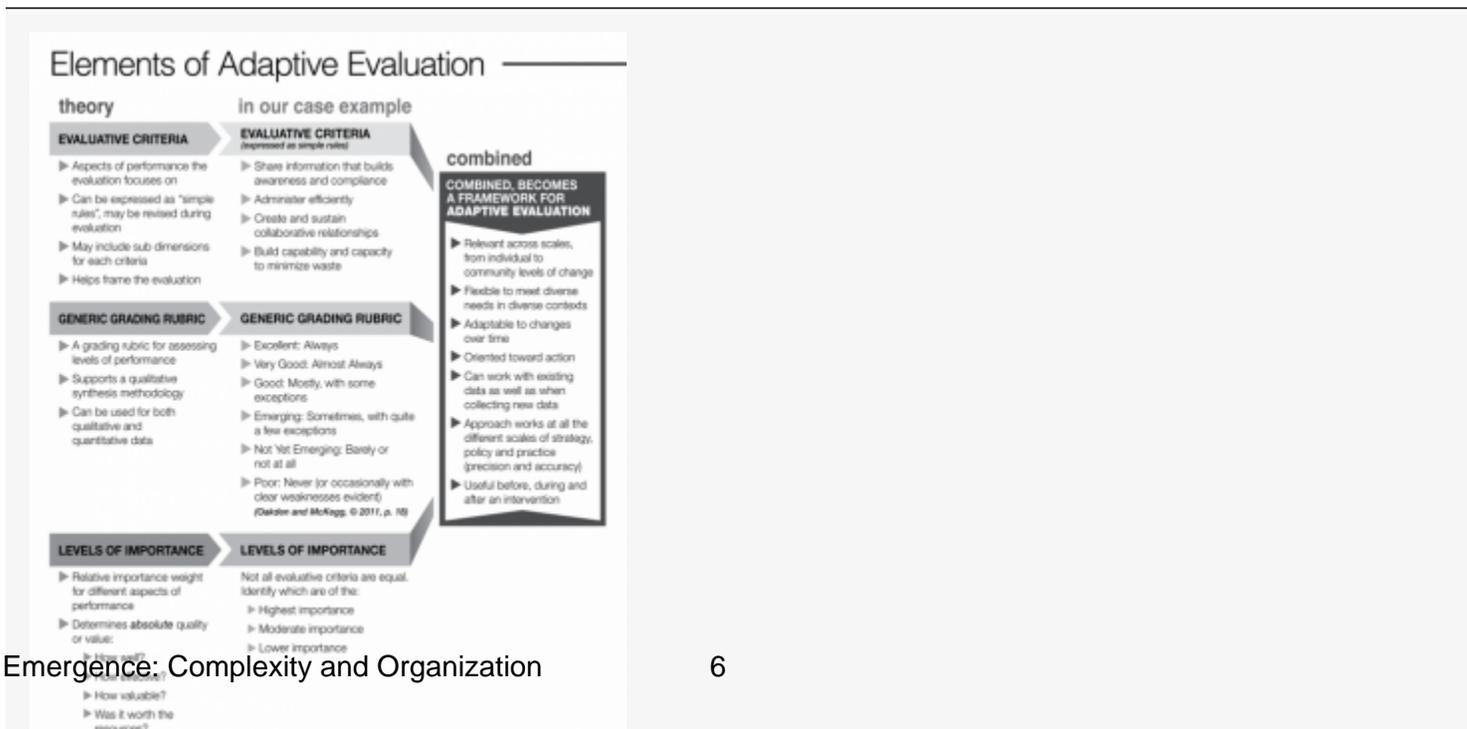
The members of the Kinnect Group, of which Judy Oakden is a member, have written several articles about using evaluation-specific methodology in practice, *Evaluative Rubrics: a Method for Surfacing Values and Improving the Credibility in Evaluation* ¹⁵, *Evaluation Rubrics: How to Ensure Transparent and Clear Assessment that Respects Diverse Lines of Evidence*¹⁶ and *Using Economic Methods Evaluatively*¹⁷. They continue to explore the power of the evaluation-specific methodology in their work in a number of ways.

What is an Adaptive Evaluation? In building on to this body of work, Glenda Eoyang and Judy Oakden have viewed the evaluation-specific methodology through the HSD lens. They have chosen to use the term *Adaptive Evaluation* to describe an approach to the evaluation-specific methodology used alongside the HSD practices that they believe appears useful in complexity. They acknowledge that Michael Quinn-Patton suggests the term Adaptive Evaluation as another term for Developmental Evaluation³².

Our approach to Adaptive Evaluation comes from combining three key components of the evaluation-specific methodology, which we believe have the potential to have great utility when evaluating in complexity. These three components are:

1. Evaluative criteria, which describe the **aspects of performance** the evaluation will focus on.
2. A generic grading rubric which describes the **levels of performance**. This rubric determines the *absolute* rather than relative quality or value¹.
3. Levels of importance, describe the **importance attributed** to each of the evaluative criteria.

The following illustration demonstrates how these three components combine to make up an Adaptive Evaluation. The theory aspect draws on both theory from evaluation^{1,18,15,14,13} and human systems dynamics^{3,8}. The practice draws from a project example^{19,20,21}, and the final column is our reflection on the benefits this brings to evaluating in complexity.



The following case study demonstrates the application of these components in a practice example.

Introduction to the case example

This section retrospectively demonstrates through a HSD lens, how an evaluation can be undertaken in complexity. In the section that follows, a case study provides an overview of how an Adaptive Evaluation approach was used to:

1. Focus the evaluation
2. Help in the management of data collection, analysis, synthesis and reporting processes
3. Be responsive to the emergent context.

The challenge

In 2010 Kinnect Group members Judy Oakden and Kate McKegg (the evaluators) assisted a Central government agency to develop an evaluation approach to identify and address any issues emerging from the introduction of a new Act, which had been passed two years before. The Act required the Central Government agency to establish and operate:

1. A waste disposal levy for waste disposal facility operators
2. Product stewardship schemes for manufacturers
3. A Waste Minimization Fund, with a portion to be paid to local authorities and a portion to be contestable
4. A Waste Advisory Board.

The Act also heralded a change in the roles and responsibilities between the Central government agency and local government (Territorial Authorities). Territorial Authorities became responsible for developing waste management and minimization plans for their regions, instead of this being undertaken at a national level.

The foundations

The evaluators were contracted to “assess how effectively the [Act] was implemented from a stakeholder perspective²⁰. This included stakeholder perceptions of barriers and enablers to implementation, emerging short-term outcomes and the impact of the changed regulatory environment on stakeholders²⁰. The Central government agency clearly articulated the intended outcomes of the Act to the evaluators using an intervention logic of the Act's implementation.

The evaluators identified the emergent context as a critical feature to be managed in this evaluation design. While some of the intended aspects of implementation had gone ahead, other aspects had not, or had only been partially implemented. In *The Strategy Process*, Mintzberg, Ghoshal and Quinn²², observe that strategy generally unfolds in an emergent manner, and that deviation from the plan may be entirely appropriate and responsive to the emerging context. This advice also can apply in an implementation setting.

Therefore, at the outset, the evaluators determined the actual level of implementation from stakeholders rather than relying on early documentation of the planned implementation as part of the scoping stage. The evaluators were also mindful that these unplanned, or unintended aspects of the implementation had the potential to produce either positive or negative outcomes²³. They planned to capture both the positive and the negative. For these reasons the evaluators recommended that evaluative criteria, rather than the goals or objectives of the Act, should be used to frame the evaluation.

Containers

As Glenda has identified earlier in this article there are a number of containers that can be relevant in a complex environment.

And that was true of this evaluation. Firstly, the evaluators established which work streams were to be included and excluded from the evaluation. For instance, the client confirmed that the establishment of the Waste Advisory Board was out of scope for the evaluation, as most stakeholders had not engaged with the Board, but most other work streams were in scope.

Another important container was the scale at which the evaluation was to be framed. Initially in developing the evaluative criteria the evaluators developed criteria for each of the different streams of work. After a brief trial, it quickly became apparent that the implementation was so complex that this approach would become too unwieldy. However, the evaluators noticed that similar patterns were emerging across evaluative criteria for different work streams. In *Coping with Chaos*, Eoyang maintains that “all organisations are fractal. They take a small set of operating principles and apply them in numerous unique situations to generate an overall pattern of behavior”²⁴. The evaluators wondered if the repetition seen in the evaluative criteria might reflect a bigger pattern of behavior overall.

Differences

A key HSD approach to uncover patterns is to “look for those ‘differences that make a difference’ in the system”³. By looking for the similarities in all these different work streams it was apparent that the range of activities broadly fell into a few categories:

1. New business systems were being developed to handle money being paid in levies
2. People applied for funding and that process had to be managed
3. Different groups from different sectors were being expected to work collaboratively together in ways they hadn’t done before
4. The Central government agency was also working in different ways with a range of stakeholders and now had both an advisory and regulatory role
5. There were new data collection, reporting and evaluation roles for a number of stakeholders.

High level overarching evaluative criteria (aspects of performance the evaluation focused on) were then developed, against which the overall implementation of the Act could be assessed. The evaluative criteria were:

- “Administrative efficiency
- Relationships — collaboration in the sector
- Good practice — building capability/ capacity (including infrastructure) across the sector
- Information, awareness and compliance.”²⁰

Exchanges

Because this was an implementation evaluation, the exchanges were the key factors to influence the success of the project. Each evaluative criterion had a number of sub criteria which provide far greater nuance for that aspect of performance. For example, for “*Administrative efficiency*” the sub criteria were:

1. “Administrative costs are in line with expectations (for Product Stewardship, Fund applications, Disposal Facility Operators and Territorial Authorities)
2. There is an appropriate balance between administrative spend and efficiency
3. [There is] a balance of projects funded
4. [There is] added value of administration to applicants — find it helpful to business, has secondary benefits
5. [There are] low levels of complaints/challenges to processes
6. [the] administrative processes are well set up, timely or robust, accurate and credible.’²⁰

Through these sub-dimensions, the different work streams or parts of the system were able to be addressed.

The Approach

The approach of Adaptive Evaluation includes six steps: 1) Develop evaluative criteria, 2) Collect data, 3) Analyze, 4) Adapt, 5) Synthesize, 6) Report. Each of the steps is described below.

Develop evaluative criteria

The evaluative criteria outlined above, were developed by the evaluators from:

1. Engaging during two sessions with a range of Ministry staff charged with policy input, implementation and operation of the Act
2. Reviewing existing documentation, working papers and background documents.

These evaluative criteria provided a strong evaluative framework which underpinned the process of evaluation of the Act overall. The client reflected that in “tight timeframes [this was a] heavy time investment for...staff — yet [the] tools [were] very efficient once [the] framework was created”¹⁹.

In addition, the evaluative criteria appeared to the evaluators to be the ‘basic rules of operation’ underpinning the implementation overall. Eoyang & Berkas observed that “the short list of Simple Rules is one mechanism that connects the parts to each other and the whole and brings the coherence of scaling to the otherwise apparently order less behavior of a CAS³. More recently Judy wondered if these overarching evaluative criteria might have been more elegantly expressed if they had been written as ‘Simple Rules’ as used in HSD work²⁴. In discussion with Glenda and Royce in 2015, the original evaluative criteria were re-expressed as Simple Rules. The table below shows the evaluative criteria developed with the key stakeholders on the left, and the corresponding Simple Rules on the right²¹. Glenda and Judy believe the version on the right could have been a more observable and a more action-oriented way to express the evaluative criteria.

Table 1

Original evaluative criteria re-expressed as ‘Simple Rules’.	
Evaluative criteria	Simple rules
Information, awareness and compliance (both in general and MfE’s performance)	Share information that builds awareness and compliance
Administrative efficiency	Administer efficiently
Relationships — collaboration in the sector	Create and sustain collaborative relationships
Good practice — building capability/capacity (including infrastructure) across the sector	Build capability and capacity to minimize waste

Having determined the focus of the evaluation, the evaluators still needed to determine how performance might be judged. One approach is a generic grading rubric for “converting descriptive data into ‘absolute’ (rather than ‘relative’) determinations of merit”¹. For this evaluation, six levels of performance were developed in conjunction with the client for the performance rating system. The evaluators reviewed and approved this performance rating schema with the client before data collection commenced.

Table 2

Performance rating system	
Rating	Description
Excellent: (Always)	Clear example of exemplary performance or great practice; no weaknesses
Very good: (Almost Always)	Very good to excellent performance on virtually all aspects; strong overall but not exemplary; no weaknesses of any real consequence
Good: (Mostly, with some exceptions)	Reasonably good performance overall; might have a few slight weaknesses but nothing serious
Emerging: (Sometimes, with quite a few exceptions)	Some evidence of performance; may be patchy; some serious but non-fatal weaknesses evident on a few aspects
Not yet emerging: (Barely or not at all)	No clear evidence of performance has yet emerged (but there is also no evidence of poor performance)
Poor: Never (Or occasionally with clear weaknesses evident)	Clear evidence of unsatisfactory functioning; consistent weaknesses across the board or serious weaknesses on crucial aspects

Collect Data

The evaluators used the evaluative criteria as a framework to both map the data to be collected and to determine which key stakeholders to interview to meet overall data collection requirements. To be credible, data collection incorporated “multiple strategies, cycle times, horizons, dimensions, informants” as recommended by Eoyang & Berkas³. This included feedback from an online survey with a wide range of different stakeholders from the different sectors and from within the Central government agency, stakeholder focus groups, in-depth interviews with key opinion former stakeholders and a range of existing administrative data from the Central government agency. The evaluators also used a Soft Systems Methodology tool, ‘rich pictures.’ With a range of stakeholders, it allowed them to better understand the “problematical situations”²⁵ the Act was addressing, how implementation of the Act was progressing and the role of different stakeholders in this process.

During data collection stakeholders told the evaluators not only about how they felt about the implementation of the Act, but how this differed from their expectations of the Act in the consultation period prior to the Act being enacted. Eoyang & Berkas recommend evaluators “capture and preserve “noise” in the system”³. This additional information provided useful insight as to why there was good awareness and understanding of the Act early in the implementation.

Analyze

As data was collected, it was recorded and analysed against the evaluative criteria in excel spreadsheets. At this stage the evaluative criteria functioned as a proxy for ‘themes’ for the data analysis. For example, a series of questions from the online survey about administrative efficiency were analysed, collated and then synthesized into a judgement of ‘good’ and the data that supported that judgement recorded. Then as other information about administrative efficiency was collected it was collated in additional columns of the spreadsheet. In this way all the information on administrative efficiency was in one place.

Adapt

A key part of an evaluation-specific methodology is determining which performance aspects are more important than others. This generally involves ethical decisions about what is valued, by whom, in what circumstances. These priorities may also shift over time. Given this evaluation was being undertaken in an emergent context, the evaluators were aware that the importance weightings of different aspects of performance might change during the evaluation. This approach aligns with an HSD suggested approach to change which suggests at times taking an infinite games approach⁸ — remaining open to difference, not setting bars for performance too early prior to making judgments of performance. This turned out to be a useful framing for this evaluation.

On this occasion, the Central government agency indicated which evaluative criteria were more important, drawing on a depth

of knowledge from engaging with the sectors. Initially client interest was in the aspects that might be linked with the introductory stage of implementation such as whether stakeholders knew about the Act and had seen information about the changes they needed to make. By the end of the evaluation, when the evaluators came to synthesize the findings, the organisation was more interested in aspects that might be linked with consolidation — for instance whether the processes were administratively efficient and the extent to which effective relationships were starting to form.

The evaluators revised the evaluation design, accommodating the changed data synthesis and reporting requirements. Because the data collected was mapped against the evaluative criteria and sub criteria, it was a relatively simple process to change the weight and focus in the synthesis and reporting for the different aspects of performance. This small cycle of Adaptive Action helped keep the evaluation on track and relevant to the client.

Synthesize

The evaluators prepared a summary of key data to share with a range of Central government agency staff charged with policy input, implementation and operation of the Act. These staff then took part in a synthesis process using the Pattern Spotting method.

The Pattern Spotting method originates from the work of Phil Capper and Bob Williams. It was originally published as CHAT Cultural-historical Activity Theory²⁶. This method is the same as a key component of HSD who describe it as ‘Pattern Spotting’⁸. The method involved five stages:

1. Stage One: Take a broad overview looking at the data overall — before getting into the detail. Ask, in general what is this data telling us, and identify the key generalizations.
2. Stage Two: Then for each generalization ask, what are the exceptions? Look to see if there are any outliers — either really excellent or poor ratings that need to be taken into account? Why might these be?
3. Stage Three: Next look for the contradictions — these might provide insights into disturbances in the system. How much should these be emphasized? Are they minor/major issues, or are they deal breakers? Are these aspects sufficiently important that it might warrant changing the rating given for general performance noted above?
4. Stage Four: Then look for things that are surprising — either because they are there, or are missing. And consider what might be learned from them?
5. Stage Five: Finally, consider what is still puzzling, and explore these puzzles rather than explain them away. Consider whether there any possible alternative explanations that might not have yet considered? What might be learned from this information?

Once the five stages were completed, a final check was made on whether the judgments seemed sensible and whether there was sufficient evidence to be credible and plausible.

Feedback from the client indicated the Pattern Spotting method made transparent to staff the process of making the evaluative judgments for this evaluation. The Pattern Spotting process also identified gaps in the data that needed further exploration prior to reporting. The Pattern Spotting session was treated as a data session in of itself as new information came to light during that session. Also the session was effective in transferring some of the key learnings to staff in a timely manner, so they could start to action key learnings prior to the report being written.

Report

The reporting was framed around the four evaluative criteria and included a dashboard that illustrated the key evaluation judgments overall. Reporting also illustrated how progress had been made for each work stream²⁰. From the reporting the internal evaluator developed a presentation for a wider internal audience. The organisation appreciated that the evaluation made evaluative judgments rather than leaving them to “figure it out”¹⁹ for themselves. In addition, the report was made public to ensure transparency to the wide range of stakeholders in a range of sectors with an interest in the findings.

Learnings from this case study

This case study illustrates the theory and practice of HSD in action, and its implications for evaluation. In particular, it illustrates the value of:

1. Adaptive Action: within the evaluation there were lots of cycles of What? So What? Now What? as the evaluators grappled with the changing context. From a client perspective, this benefited them as it allowed “more flexibility in working collaboratively”¹⁹.
2. Simple Rules were expressed approximately in this evaluation. We have seen that they could have been expressed better. However, the benefits were still accrued from the client perspective and it was evident that the evaluative criteria:
 - “Framed the evaluation differently from traditional linear thinking — captures complex dynamics and inter-linkages...
 - Provided a broad-brush framework for evaluating different activities within one initiative, and captured ‘messy’ hard-to-measure dimensions
 - Enabled effective and aligned mixed methods data collection, synthesis and actionable reporting.”¹⁹
3. Standing in inquiry: the evaluators stood in inquiry throughout the evaluation and were not thrown by the emerging context. This is not always a comfortable space. It was constantly a case of ‘feeling our way’. This also was greatly assisted by having a client who saw the benefit in this way of working, and who wanted others to “see the ‘big picture’ connections across silos.”¹⁹.

Conclusion

This paper updates Glenda Eoyang and Tom Berkas³ thinking on evaluating in complexity. In this paper, synergy has emerged between complexity theory, through the lens of HSD and evaluation practice, through a case study of a complex program of social change. What emerges at this generative intersection is an evaluation approach, *Adaptive Evaluation*, that is simple, robust and flexible enough to meet the demands of twenty-first century social change.

On reflection key benefits of the Adaptive Evaluation approach appear to be that it:

1. Can be used in evaluation at many stages ranging from formative to summative evaluations,
2. Provides an evaluation design that is flexible enough to accommodate diverse needs in diverse contexts
3. Can cope with multiple and unpredictable data sources
4. Can accommodate change over time without losing continuity
5. Makes space for useful stakeholder involvement throughout the process.

Clients recognize Adaptive Evaluation is oriented towards action. The rich evidence base is seen as credible, which builds client confidence to buy into and inform future action¹⁹. While untested in this specific evaluation case study, we believe Adaptive Evaluation also has the potential to work well in community settings and in cross-cultural settings²⁷.

What are the limitations of this kind of approach? Adaptive Evaluation is an approach for use in complexity, so it will not thrive in situations where:

1. Ambiguity and emergence are not tolerated, for instance if procurement requires a tightly scoped evaluation plan from the outset and assumes the evaluation will go ahead with minimal change, even over a longer time-frame
2. The client and those who might use the evaluation do not want to engage with the evaluators (seeing this engagement as reducing the rigor of the evaluation)
3. The evaluators are uncomfortable standing in inquiry. In our experience evaluators have to be prepared to let go of control and certainty and instead be curious and non-judgmental as they respond to an unfolding situation.

By using an HSD lens to examine this evaluation practice we believe we have made sense of and realized the importance of

some of the aspects of Adaptive Evaluation that were not initially apparent. We hope this praxis example is of use to other evaluators looking for effective ways to design and undertake evaluation in complexity.

Acknowledgements

The authors would like to acknowledge and thank the Ministry for the Environment, Wellington, New Zealand for allowing us to use the 'Waste Minimisation Act implementation: evaluation of stakeholder perceptions' project as the case study for this article.

References

1. Davidson, E. J. (2005). *Evaluation methodology basics: the nuts and bolts of sound evaluation*. Thousand Oaks: Sage. ISBN: 978-0-471-98606-5
2. Patton, M. Q. (2011). *Developmental evaluation: Applying complexity concepts to enhance innovation and use*. New York: Guildford Press. ISBN: 978-1606238721
3. Eoyang, G. H., & Berkas, T. H. (1999). Evaluating performance in a complex adaptive system (CAS). In M. Lissack, & H. Gunz, *Managing complexity in organizations: a view in many directions* (pp. 313-335). Westport, CT: Quorum Books. ISBN: 978-1567202854
4. Kuhn, T. S. (1996). *The structure of scientific revolution*. Chicago: Chicago: University of Chicago Press. ISBN: 978-0226458083
5. Williams, B., & Imam, I. (2007). *Systems concepts in evaluation: An expert anthology*. Point Reyes, CA: EdgePress of Inverness. ISBN: 978-0918528216
6. Reynolds, M. (2015). (Breaking) The Iron Triangle of evaluation. *IDS Bulletin* 46(1), 71-86. DOI:10.1145/37401.37406
7. Boulton, J. G., Allen, P. M., & Bowman, C. (2015). *Embracing complexity: Strategic perspectives for an age of turbulence*. Oxford: Oxford University Press. ISBN: 978-0199565269
8. Eoyang, G., & Holladay, R. (2013). *Adaptive Action: Leveraging uncertainty in your organization*. Stanford: Stanford University Press. ISBN: 978-0804787116
9. Reynolds, M. (2014). Equity-focused developmental evaluation using critical systems thinking. *Evaluation — The International Journal of Theory, Research and Practice*, 20(1) 75—95. DOI: 10.1177/1356389013516054
10. Wilson-Grau, R. (2016). *Outcome Harvesting*. Retrieved September 11, 2016, from <http://www.outcomeharvesting.net/>
REFERENCE LINK
11. Parsons, B. (2009). Evaluative Inquiry for Complex Times. *OD Practitioner*, 41, 1. Pp 44-49. DOI: 10.1.1.454.7741
12. Eoyang, G. H. (2006). Human systems dynamics: Complexity based approach to a complex evaluation. In B. Williams, & I. Imam, *Systems concepts in evaluation: An expert anthology*. Point Reyes: American Evaluation Association.
13. Scriven, M. (2012). The Logic of Valuing. *New Directions for Evaluation*, 133, 17-28. doi:<http://dx.doi.org/10.1002/ev.20003>
14. Nunns, H., Peace, R., & Witten, K. (2015). Evaluative reasoning in public sector evaluation in Aotearoa New Zealand: how are we doing? *Evaluation Matters - He Take To Te Aromatawai*, *Journal of the Aotearoa New Zealand Evaluation Association*. DOI: dx.doi.org/10.18296/em.0007
15. King, J., McKegg, K., Oakden, J., & Wehipeihana, N. (2013). Evaluative rubrics: a method for surfacing values and improving the credibility of evaluation. *Journal of MultiDisciplinary Evaluation*, 9(21). ISBN 978-0-473-27326-2
16. Oakden, J. (2013). *Evaluation rubrics: how to ensure transparent and clear assessment that respects diverse lines of evidence*. Melbourne: Better Evaluation. Retrieved from <http://betterevaluation.org/en/resource/example/rubrics-oakden>
REFERENCE LINK
17. King, J. (2016). Using economic methods evaluatively. *American Journal of Evaluation* , 1-13. DOI: 10.1177/1098214016641211
18. Fournier, D. (1995). Establishing evaluative conclusions: A distinction between general and working logic. *New Directions for Evaluation*, 68, 15-32. DOI: 10.1002/ev.1017
19. Oakden, J. and Bear, C. (2011). "Managing complexity in evaluation," presented at the Aotearoa New Zealand Evaluation Association (ANZEA) Conference, 8-10 August, Wellington.
20. Oakden, J., & McKegg, K. (2011). *Waste Minimisation Act implementation: evaluation of stakeholder perceptions*.

Wellington : Kinnect Group. Retrieved from <http://www.mfe.govt.nz/publications/waste/waste-act-stakeholder-perception-report-final/index.html>
REFERENCE LINK

21. Oakden, J., & Eoyang, G. (2015). Evaluation rubrics look easy but can be hard to do well: lessons from the field. Presented at the American Evaluation Association Conference, November 2015. Chicago.
22. Mintzberg, H., Ghoshal, S., & Quinn, J. B. (1998). *The Strategy Process*. Upper Saddle River: Prentice Hall. ISBN 013675984X
23. Scriven, M. (1991). Prose and Cons about Goal-Free Evaluation. *Evaluation Practice*, 12(1), 55-76. DOI: [http://dx.doi.org/10.1016/0886-1633\(91\)90024-R](http://dx.doi.org/10.1016/0886-1633(91)90024-R)
24. Eoyang, G. H. (1997). *Coping with chaos: seven simple tools*. Cheyenne: Lagumo. ISBN: 978-1878117151
25. Checkland, P. (1999). *Systems thinking, systems practice: Includes a 30-year retrospective*. Chichester: Wiley. ISBN: 978-0-471-98606-5
26. Capper, P., & Williams, B. (2004). Enhancing evaluation using systems concepts CHAT. Presented at the American Evaluation Association Conference, November, 2004, Atlanta. Retrieved from http://www.bobwilliams.co.nz/Systems_Resources_files/activity.pdf
REFERENCE LINK
27. Stone-Jovicich, S. (2015). To rubrics or not to rubrics: An experience using rubrics for monitoring, evaluating and learning in a complex project. Canberra: Commonwealth Scientific and Industrial Research Organisation. DOI: 10.13140/RG.2.1.3370.8565
REFERENCE LINK
28. Holladay, R. (2005). Simple Rules: Organizational DNA. *OD Practitioner: Journal of the Organization Development Network*, 36(3).
29. Human Systems Dynamics Institute. (2012). Simple Rules. Retrieved from <http://www.hsdinstitute.org/resources/simple-rules.html>
REFERENCE LINK
30. Human Systems Dynamics Institute. (2012). Pattern Spotters. Retrieved from <http://www.hsdinstitute.org/resources/pattern-spotters.html>
REFERENCE LINK
31. Julnes, G. (2012). Developing policies to support valuing in the public interest. *New Directions for Evaluation*, 133, 109-129. DOI: 10.1002/ev.20012
32. Patton, M.Q., McKegg, K. and Wehipeihana, N. (eds.) (2015). *Developmental Evaluation Exemplars: Principles in Practice*, ISBN 9781462522965.