Evaluating Complex and Complicated Programs: Issues, Approaches, Implementation and Implications

Australasian Evaluation Society Conference
Darwin
September 2006

*Lessons from the Evaluation of the Stronger Families and Communities Strategy 2000-2004**

Sue Funnell and Patricia Rogers

*** Funded by the then Australian Government Department of Family and Community Services**

Overview

Ways of thinking about simple, complicated and complex

Implications for evaluating complicated and complex programs

Evaluation strategies for addressing some of these implications

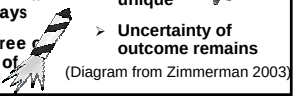
Application of these strategies to Evaluation of the Stronger Families and Communities Strategy 2000-2004

Implications for future evaluation of complicated and complex programs

Ways of thinking about complicated/complex

- Synonyms – 'complex' sounding more scientific or problematic
- Various aspects of complexity theory
- Detailed but predictable (complicated) cf emergent and unpredictable (complex) (Glouberman and Zimmerman, 2002)

Simple Following a Recipe	***Complicated*** A Rocket to the Moon	***Complex*** Raising a Child
• **The recipe is essential** • **Recipes are tested to assure replicability of later efforts** • **No particular expertise; knowing how to cook increases success** • **Recipes produce standard products** • **Certainty of same results every time**	➢ **Formulae are critical and necessary** ➢ **Sending one rocket increases assurance that next will be ok** ➢ **High level of expertise in many specialized fields + coordination** ➢ **Rockets similar in critical ways** ➢ **High degree of certainty of outcome**	➢ **Formulae have only a limited application** ➢ **Raising one child gives no assurance of success with the next** ➢ **Expertise can help but is not sufficient; relationships are key** ➢ **Every child is unique** ➢ **Uncertainty of outcome remains**



(Diagram from Zimmerman 2003)

Program characteristics

Simple programs	**Complicated programs**
Single organisation (or clear contractual relationships)	**Multiple organisations** working together in network governance
One program even if implemented at different sites	**Diverse projects** implemented under a single funding program or policy
Single causal strand to the program theory If A then B	**Multiple simultaneous causal strands** - If A and C (or A in context C) then B
Single causal strand to the program theory If A then B	**Multiple alternative causal strands** – if A or C or D then B
Simple programs	**Complex programs**
Linear causality	**Recursive causality** – a small initial effect may lead to a large ultimate effect through a reinforcing loop or critical tipping point
Specific pre-defined outcomes	**Emergent specific outcomes** in response to opportunities

Implications of program characteristics

Complicated or complex programs	**Implications for evaluation**
Multiple organisations	Need to negotiate agreement about evaluation parameters and processes
Diverse projects	Synthesis will not be simply arithmetic
Multiple simultaneous causal strands	Evaluation should both document and support the different causal strands that programs need to achieve, not just one To inform replication of an effective program evaluation may need to understand the context that supports it
Multiple alternative causal strands	To inform replication of an effective program evaluation may need to understand the context that supports it
Recursive causality	Since a small initial effect may lead to a large ultimate effect through a reinforcing loop or critical tipping point, evaluation needs measurement over time not at one point
Emergent outcomes	Specific measures may not be able to be developed in advance, making pre- and post-comparisons difficult

Addressing some of these challenges

Complicated or complex programs	Strategies for evaluation
Diverse projects	
Multiple simultaneous causal strands	
Multiple alternative causal strands	
Recursive causality	
Emergent outcomes	

Addressing some of these challenges

Complicated or complex programs	Strategies for evaluation
Diverse projects	**Overarching program logic** that can be adapted specifically for diverse projects **Non-arithmetic synthesis methods** for comparing diverse projects with diverse evaluation evidence
Multiple simultaneous causal strands	**INUS analysis** (exploring factors that are necessary but not sufficient to produce the intended outcome)
Multiple alternative causal strands	**INUS analysis** (exploring factors that are sufficient but not necessary to produce the intended outcome), particularly the context within which causal paths work
Recursive causality	**Non-linear program logic models** showing and documenting iterative development over time
Emergent outcomes	**Overarching program logic** that can be adapted over time to accommodate specific emergent outcomes

Evaluation case

- Stronger Families and Communities Strategy 2000-2004
- Australian Government (then Department of Family and Community Services) initiative to help build family and community capacity to deal with challenges and take advantage of opportunities
- Special focus on those at risk of social, economic and geographic isolation
- Main focus of the evaluation was 635 projects funded under seven community-based linked initiatives

(Information about the 2005-2009 Strategy is available at http://www.facsia.gov.au/internet/facsinternet.nsf/aboutfacs/programs/sfsc-sfcs.htm)

Main focus of the evaluation

Community-focused initiatives:

- Potential Leaders in Local Communities (144 projects $19.7m in funding)
- Local Solutions to Local Problems (207 projects $8.7m)
- National Skills Development for Volunteers (26 projects $4.3m)
- Can Do Community (14 projects $0.96m)

Family-focused initiatives:

- Early Intervention, Parenting & Family Relationship Support (184 projects $26.8m)
- Stronger Families Fund (49 projects $18.2m)
- Early Childhood Initiative (Strategy component) (11 projects $1.2m)

Contributors to the evaluation

Initial Evaluation Framework developed by SuccessWorks

Evaluation Framework finalised and implemented between 2002-2005 by a consortium

- led by CIRCLE at RMIT University (Collaborative Institute for Research Consulting and Learning in Evaluation, Royal Melbourne Institute of Technology) with
- Performance Improvement (Sue Funnell)
- Bearing Point (Australia)
- John Scougall
- Other researchers

Complicated and complex aspects of SFCS

Aspects	SFCS features
Diverse projects	635 projects that varied in terms of: • specific objectives • activities • duration • target groups • scale of projects ($1K to over $1M) • starting points for families, communities and organisations • accessibility/remoteness
Multiple simultaneous causal strands	
Multiple alternative causal strands	
Recursive causality	
Emergent outcomes	

Complicated and complex aspects of SFCS

Aspects	SFCS features
Diverse projects	635 varied projects
Multiple simultaneous causal strands	
Multiple alternative causal strands	
Recursive causality	
Emergent outcomes	

Complicated and complex aspects of SFCS

Aspects	SFCS features
Diverse projects	635 varied projects
Multiple simultaneous causal strands	Many different factors needed for success - no 'silver bullet' Contribution of other projects before, during and after Strategy projects
Multiple alternative causal strands	
Recursive causality	
Emergent outcomes	

Complicated and complex aspects of SFCS

Aspects	SFCS features
Diverse projects	635 varied projects
Multiple simultaneous causal strands	Many different factors needed for success - no 'silver bullet' Contribution of other projects before, during and after Strategy projects
Multiple alternative causal strands	Alternative ways of achieving success – not a simple 'no treatment' counter-factual Some causal strands only effective in particular contexts
Recursive causality	
Emergent outcomes	

Complicated and complex aspects of SFCS

Aspects	SFCS features
Diverse projects	635 varied projects
Multiple simultaneous causal strands	Many different factors needed for success - no 'silver bullet' Contribution of other projects before, during and after Strategy projects
Multiple alternative causal strands	Alternative ways of achieving success – not a simple 'no treatment' counter-factual Some causal strands only effective in particular contexts
Recursive causality	Initial success or failure can have a consequential effect on further investment by all parties and opportunities
Emergent outcomes	

Complicated and complex aspects of SFCS

Aspects	SFCS features
Diverse projects	635 varied projects
Multiple simultaneous causal strands	Many different factors needed for success - no 'silver bullet' Contribution of other projects before, during and after Strategy projects
Multiple alternative causal strands	Alternative ways of achieving success – not a simple 'no treatment' counter-factual Some causal strands only effective in particular contexts
Recursive causality	Initial success or failure can have a consequential effect on further investment by all parties and opportunities
Emergent outcomes	Specific objectives were evolving and responsive to emerging needs, opportunities, solutions Many projects focused on capacity building in response to community identified issues Specific objectives were often locally determined

Strategies applied

1. **Overarching program logic** that can be adapted for diverse projects and emergent outcomes
2. **Non-linear program logic** showing and documenting iterative development over time
3. **Emergent evaluation design** to address emergent outcomes and explore complicated causal paths
4. **Non-arithmetic methods for synthesis** of diverse evidence
5. **INUS Analysis** to explore factors that may be necessary but not sufficient or sufficient but not necessary, particularly the context within which they work

1. Overarching program logic

- Provided a **conceptual framework** to incorporate the diversity of project objectives and emergent objectives
- Provided an **analytical framework** for coding specific outcomes
- SMART objectives for the program as a whole were inappropriate (although they may have been appropriate for some individual projects)

Common Hierarchy of Outcomes

7. Stronger Families and Communities

⇧

6. An environment where communities participate in and drive their own solutions to strengthen their families and communities

⇧

5. Family and community trust / resilience / adaptability

⇧

4. Demonstration / application of greater understanding, skills and capacity

⇧

3. Greater choice, understanding, skills, capacity for initiative

⇧

2. Greater awareness, development of partnerships

⇧

1. Participation, enhanced trust

Adaptations and extensions of the overarching outcomes hierarchy

- **Adaptations**:
  - To types of projects e.g. capacity building
  - To specific projects e.g. project working with young parents, a community garden.
  - Potential for adaptation across projects not fully met
- **Extensions**:
  - Success criteria and definitions for each level of the outcomes hierarchy to address range of types of specific objectives
  - Factors that affect success – program and non-program
  - For each level of outcome, specific evaluation questions/types of information required, methods of data collection, FaCS performance indicators

2. Non-linear program logic

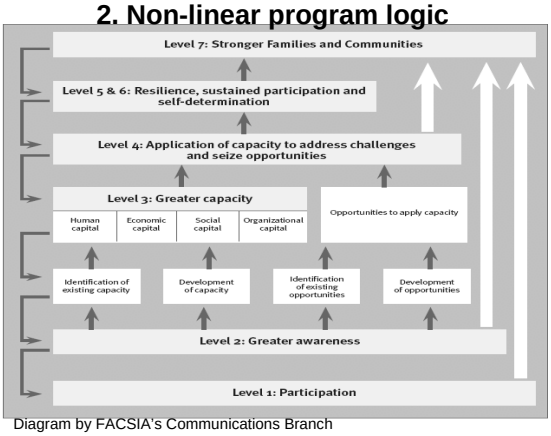


Diagram by FACSIA's Communications Branch

Use of non-linear program logic

- Search for **iterative progress** – eg cycles of developing and then using community capacity – in developments since project ended and/or in subsequent project
- Search for outcomes related to stronger families and communities (for example, improvements in physical wellbeing) that **resulted directly** from lower-level outcomes, not through the entire sequence of outcomes (ie capacity building)

3. Emergent evaluation design

Level 1	Data potentially available for all projects – activities, resources used, outcomes in terms of the common program logic
Level 2	Issue-focused papers that drew on research evidence and policy frameworks and illustrative data from a purposeful sample of Strategy projects
Level 3	Case studies of specific Strategy projects, communities and initiatives
Level 4	Synthesis of all data sources

Level 1 data

Questionnaires
- Initial (soon after commencement) – project development processes
- Final (near completion) – activities, outcomes

Performance Indicators and Performance Information
- Progress reporting
- Final reporting

Project Reports
- Progress reports
- Final reports

Level 2 issue papers

- Focused on issues of significance across the Strategy:
  - Networks and partnerships
  - Community capacity building
  - Early intervention, particularly in early childhood
  - Sustainability and legacy
  - Economic and social participation
  - Service integration and co-ordination
  - Evidence-based policy and practice
- Drew on research and policy literature with illustrations from Strategy projects
- Set out key ideas and operationalised them for more systematic empirical investigation in other parts of the evaluation

Level 3 case studies

- Specific focus of these decided progressively during the evaluation
  - Individual projects
    - Gilles Plains Community Garden
    - Indigenous capacity building project
    - Indigenous integrated family strengthening project
    - Indigenous community leadership development project
  - Specific funding initiatives
    - Early Intervention Initiative
    - Stronger Families Fund initiative
    - Potential Leaders in Local Communities initiative
  - Other case studies
    - Mandurah targeted region
    - Lessons Learnt about Strengthening Indigenous Families and Communities: What Works?
    - Sustainability and legacyof projects
    - Qualitative cost-benefit analysis
    - Implementation of the Strategy across States and Territories
- Drew on detailed analysis of existing evidence together with some additional data collection

4. Non-arithmetic methods for synthesis

- Synthesis not aggregation of different types of data
- Looked at the relationships between types of data e.g. between qualitative data in project reports and quantitative questionnaire data
- Realist perspective meant looking for negative examples not just measures of central tendency – what does/doesn't work for whom under what circumstances

What were the preconditions for undertaking qualitative synthesis?

If we wanted to:

- **look for relationships and patterns**, then we needed to make the data from different sources manageable, given diversity
- **be confident about the different sources of data we were using**, then we needed to assess the quality of data before using it
- **judge how successful projects had been**, then we needed criteria for judging success
- **see what types of results projects and the program as a whole were achieving**, then we needed to preserve an outcomes focus

What did we do?

We used various types of coding that would enable us to look at relationships.

Coding included:

1. Assessing the **quality of evidence** provided
2. Rating the **global success** of the project
3. Classifying the **types of outcomes** achieved
4. **Additional coding** of projects funded by Early Intervention Initiative – use of level 2 issues paper to design & conduct level 3 case study

Coding of global success and types of outcomes took into account the results of coding quality of evidence.

Coding 1: Quality of evidence

Verifiable

Plausible

No or negligible evidence

Coding 1: Quality of evidence

Verifiable

- both plausible and relatively easily verifiable e.g. feedback had been documented and in principle could be requested; surveys, measures and /or other records had been used
- e.g. a recorded response from a participant in an interview or questionnaire, direct quotation.
- Some involved strong research designs such as those involving measurement over several occasions (before and after an intervention) and with comparison groups using standardised instruments but these types of study were rare and in most cases neither appropriate nor feasible given the nature of the projects and the populations they were serving.

Coding1: Quality of evidence

Plausible

- The reports included plausible claims concerning specific outcomes for specific individuals, groups or the community as a whole - that participants had gone on to do particular things as a result of what they had learnt and there was a clear logic to the connection, that they had articulated what they had learnt to the project officer, that others had observed changes in behaviour.
- e.g. a program that had marginalised youths working with older mentors to build a ramp for people with disability reported that the young people had gained skills and improved self-esteem through making a contribution.

Coding1: Quality of evidence

No or negligible evidence

- The outcome is simply claimed, typically in general terms without examples or other supporting evidence

  e.g. the community is now more cohesive.

Different approaches for coding quality of data

Projects with final reports

done for each outcome level claimed, based on a combination of final report, final questionnaire and FaCS PI data)

Projects without final reports

but with final questionnaires, with / without PI data – global assessment of quality of evidence

Coding 2: Global success of projects

Coding categories were:

- Outstanding
- Generally successful
- Mixed or moderate success
- Little or no success
- Unknown (insufficient evidence)

All categories had descriptors

To illustrate the criteria, the report included descriptions of projects at different levels of success and different size budgets, and explained why they had been coded thus

How were global ratings of success used?

- **Looking for patterns** through cross breaks of projects at different levels of success by various items in final questionnaire, and use of multiple regression to identify predictors of global success
- **Looked at exceptions to patterns** by returning to final reports to dig deeper and doing further qualitative analysis and cross breaks.

  Examples:

  1. role of the auspice
  2. SFCS principles e.g. evidence based approaches
  3. types of target groups
  4. different approaches to home visiting

Coding 3: Outcomes achieved

- Coding acknowledged that there were diverse specific outcomes but they could be analysed under the umbrella of the SFCS outcomes
- Initial coding by projects replaced by coding by evaluation team
- It was easier to find verifiable evidence for lower level than for higher level outcomes
- Causality more difficult to demonstrate for higher levels
- Process provided many specific examples of outcome levels in the SFCS Outcomes Hierarchy for inclusion in the report
- Coded outcomes were aggregated to show overall success of SFCS in relation to its Outcomes Hierarchy.

Coding 4: Effective Early Intervention

- Analysis of 44 EI projects with final reports available in 2004 with respect to eight characteristics of effective early intervention projects identified in level 2 issue paper.
- Many variables and descriptors for each characteristic
- Generated hypotheses for later exploration with more EI projects as part of case study of all Early Intervention projects
- Information from relevant items in the final questionnaires for which responses were received for 146 of the 195 EI projects.
- Illustrates use of level 2 data (Issues paper on EI) to formulate hypotheses for level 3 case study (on EI) which drew on level 1 data (questionnaires, final reports, performance indicator data)

Lessons learnt about methods used for qualitative synthesis

- Intrinsically conceptually complex analysis with much judgement required
- Iterative processes and revisiting of data required
- Resource intensive and have to know when to stop but also have to be curious and not stop too soon – look for exceptions
- Needed good and easily accessible relational data base which we had courtesy of Bearing Point (drop downs on variables, item numbers, project cross referencing between different data sets)
- Much cross checking and discussion is needed to improve reliability of coding

5. INUS analysis

- An INUS condition: an Insufficient but Necessary part of a causal package that itself is Unnecessary but Sufficient to cause the result. (Mackie, 1965) – eg short circuit causing a fire
- Not looking for 'what works' but '**what works for whom, in what circumstances and how**?"
- Searching for and explaining **negative examples** (Miles and Huberman, 1984) – those where the identified cause was not sufficient (outcome was not achieved) or not necessary (outcome was achieved in another way)

Example of INUS analysis

- 'Effective support from Auspice organisation' was identified as a contributing factor to effective EI projects
- Searched for, and explained, negative examples of projects
  - With this support that *had not* been effective (ie **insufficient**)
  - Without this support that *had* been effective (ie **not necessary**)

Implications for future evaluation of complicated and complex programs

- Evaluations that seek simple answers about "what works" are unlikely to accurately reflect the multiple causal factors and pathways which exist and may therefore not be so useful for replication or evidence-based policy
- Emergent programs may need an element of emergent evaluation design and measures
- Program logic may need to be used clearly as a heuristic not an implementation blueprint

Further information

- Glouberman and Zimmerman's simple/complicated/complex typology
  http://www.acgme.org/outcome/PowerPoint/Zimmerman.ppt
- Mackie's INUS analysis
  Mackie, J.L. "Causes and Conditional' American Philosophical Quarterly 2 (1965), 245-65.
- SFCS 2000-2004
  http://www.facs.gov.au/internet/facsinternet.nsf/aboutfacs/programs/sfsc-sfcs2000-2004.htm
- SFCS 2005-2009
  http://www.facsia.gov.au/internet/facsinternet.nsf/aboutfacs/programs/sfsc-sfcs.htm