Qualitative methods in regional program evaluation

Vanclay, Francis

Published in:
Handbook of Research Methods and Applications in Economic Geography

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2015

25 Qualitative methods in regional program evaluation: an examination of the story-based approach*
Frank Vanclay

INTRODUCTION: THE NEED FOR QUALITATIVE EVALUATION

The attempt to identify what works and why are perennial questions for evaluators, program and project managers, funding agencies, policy makers and economic geographers interested in regional development (Greene, 2000; Feller, 2007). Policies, programs, plans and projects (hereafter all 'programs' for convenience) all start with good intent, often with long-term and usually over-optimistic goals. An important issue is how to assess the success of these programs during their life, often before their goals have been fully achieved. Thus some sense of interim performance is needed – to provide feedback to fine-tune the program; to determine whether subsequent tranche payments should be made; and to assist in decision making about whether similar programs, or projects within these programs, should be funded.

Evaluation in such circumstances is complex. How can the achievement of goals be assessed if the goals are long term? Evaluation can not wait years to determine whether a program has been successful – answers are needed now to support decision making. Thus evaluation needs to consider carefully the program logic, whether interim steps have been achieved and whether there are signs that longer-term objectives and goals are likely to be achieved. But this is not straightforward. All programs, especially long-term ones, should incorporate a degree of adaptive management or reflexivity, allowing them to respond to feedback along the way. Final success, therefore, is not just whether the original plan was correct, but the extent to which a program has effective monitoring and is capable of adapting to feedback along the way. Depending on the context, it may also be the case that external factors have changed and the original goals and/or program logic need to be revised to accommodate the changed external circumstances. Any program seeking to contribute to high-level goals such as enhanced community wellbeing, social sustainability, regional development potential, innovativeness and so on is likely to be affected by a changing context. Therefore a key factor for success (and thus for evaluation) is the ability of the program to be responsive to change.

It is increasingly evident that standard *ex-post* quantitative evaluation techniques are not adequate to deal with these matters. Standard approaches to evaluation can not deal with long lag times; they can not cope with multi-causality; and they can not cope with a changing operating environment. Qualitative methods, on the other hand, offer a way of collecting evidence about the performance of a program or project. They also enable the collection of feedback to assist in modifying the program. They can work in tandem with quantitative indicators, playing a complementary but different, although equally

important, role. This chapter explores issues associated with the evaluation of regional development programs, and, using an example from Australia, highlights how qualitative methods can be used to assess the performance of regional development programs.

ISSUES FACING PROGRAM EVALUATION

Many programs facilitate the provision of a broad range of social benefits that were not necessarily the core purpose of the program, and frequently there can be many other unanticipated spin-off benefits as well. Collectively these may contribute significantly to the perceived success of the program, especially by program beneficiaries. Should evaluation consider the success of a program on the basis of unintended consequences? At face value, many key decision makers might say no. But, on the other hand, it is unlikely that a program would be considered successful if it caused a lot of unanticipated harm in addition to still achieving its narrow goals. Evaluation therefore must take an holistic approach considering the potential for good as well as the potential for harm, and it needs to consider the unanticipated consequences as much as the intended goals (Esteves and Vanclay, 2009; João et al., 2011; Vanclay, 2003).

Another issue is that an evaluation can not simply measure whether goals (i.e. desired results) were achieved. If so, how would the evaluation establish causality? Could the observed change have been the result of other things occurring at the same time? What if there were underlying trends in a community anyway? The concept of 'baseline' should not be regarded as a single data point fixed in time where the analytical comparison is time $(x + 1)$ against time (x); but rather a comparison at a point in time against what would have happened without the program (Esteves et al., 2012). The baseline is thus the trajectory line (not point) of expected trending without the program. In European policy circles, this is called the 'counterfactual' (EVALSED, 2009) – a term borrowed from psychology where it has a slightly different meaning, a mental representation or image of an alternative trajectory, past or future (Roese, 1997). Thus programs can still be regarded as 'successful' if an indicator at a future time is worse than it was at the beginning, providing that there is a reasonable analysis that there were other changes taking place such that the program made the community better off than it would have been without the program.

In considering a wide range of outcomes, and with the realization that many of the broader social benefits of programs are subjective, the old adage normally attributed to Albert Einstein that 'not everything that counts can be counted' becomes important (Vanclay, 2012). Particularly in cases of the enhanced wellbeing type of programs referred to above, the additional benefits may be in terms of an improvement in how people feel about where they live and their lives in general, about how they feel about the future of their community, and about how different groups in a community cooperate or at least get on with each other. While not necessarily impossible to measure, these high-level goals are difficult to measure, and are not normally included in routinely collected data collection processes.

The issue of high-level broader social benefits raises the question of attribution. How can the evaluator know whether an observed effect was due to the program? A short and simple answer might be that they can not. A more complex answer questions whether

simplistic assumptions of direct cause and effect are appropriate. Big programs with high-level outcomes do not have simple cause–effect relationships; they have complex interconnecting multicausal linkages. A deeper understanding of the nexus of these relationships is needed. Such systems are dynamic, are mediated by iterative feedback processes, are confounded by inhibiting and enabling mechanisms, and are potentially affected by catalytic relationships (including nonlinear and exponential effects) between system elements.

It is important to realize that these debates have existed in the field of evaluation for decades (see Greene, 2000 for a discussion on the purpose and history of evaluation). While some evaluators have attempted to persist with ever-improving and ever-more sophisticated empirical quantitative techniques (Leeuw and Vaessen, 2009), many other evaluation experts fundamentally disagree that such methods can address the complexity of the programs being considered (Guijt et al., 2011). Instead, they advocate the use of robust qualitative measures, arguing that qualitative methods are more valid, give better information, are more efficient, include the potential for unanticipated factors to be included, and address causality.

A final argument in favor of qualitative methods (especially story-based approaches) is that they can yield powerful stories that are not only useful for media reports, but are often frequently preferred by politicians and other decision makers (Denning, 2007; Kurtz, 2009; Mayne, 2004). It is an illusion of scientists that hard data are the only convincing evidence. As Benjamin Disraeli (or at least Mark Twain) implied many years ago with the now famous 'lies, dammed lies and statistics' aphorism, a statistic (data, evidence) is only as accurate as the reliability of the processes used to collect it and the extent to which it faithfully represents reality (its validity). But reliability, validity and significance, the once all-important cry of quantitative social researchers, have now been replaced by other criteria. With so much data, evidence, information and knowledge everywhere, the key concerns of the users of information are no longer the old ideal concerns of purist statisticians, but the pragmatic considerations of salience, credibility and legitimacy (Cash et al., 2003). Users of information want to know: 'Is it relevant information?'; 'Is it useful information?'; and 'Do I believe it?' – which is partly based on its credibility to them as individuals (in other words, is consistent with their worldview) and partly on the extent to which they trust the source of that information. Very often, a story conveys this information much more effectively (i.e. convincingly) than other forms of evidence (Denning, 2007; Fisher, 1989; Kurtz, 2009; Sandelowski, 1991; Shaw et al., 1998).

A NOTE ON TERMINOLOGY: OUTPUTS AND OUTCOMES

Program logic refers to the perceived causal understanding of how the different components of a program (specifically inputs and activities) work together to produce outputs and outcomes. Outcomes are generally conceived as being immediate, intermediate and ultimate (Mayne, 2004). Program logic 'captures the rationale behind a program, probing and outlining the anticipated cause-and-effect relationships between program activities, outputs, intermediate outcomes and longer-term desired outcomes. A program logic is usually represented as a diagram or matrix that shows a series of expected consequences, not just a sequence of events' (Roughley, 2009, p. 7). Mayne (1999) presents a good model illustrating the complexity of program logic (see Figure 25.1).
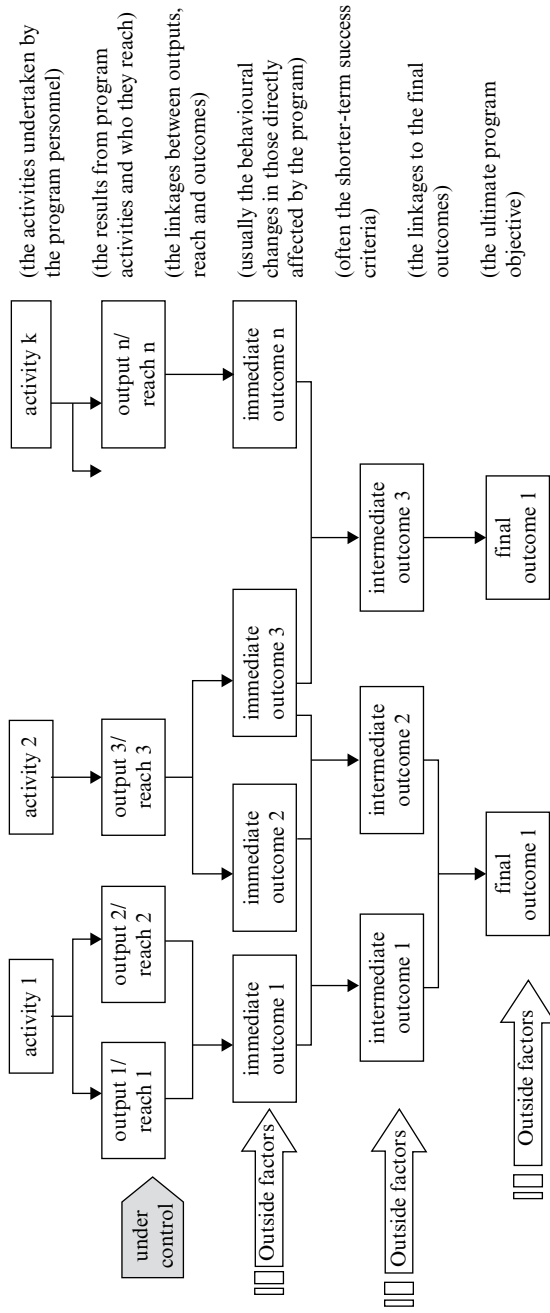
Source: Mayne (1999, p. 9).

Figure 25.1   A generic depiction of program logic

Another significant concept in the field of evaluation (especially in the area of agricultural extension) is Bennett's hierarchy (Bennett, 1975). Bennett championed the phrase 'up the hierarchy' as a reaction to the excessive attention being given to inputs (e.g. dollars spent, hours consumed), activities (e.g. numbers of workshops held) and people involvement (e.g. numbers of people attending), arguing that more attention needed to be given to a range of higher-order considerations. Bennett's model went as follows:

1. inputs (resources expended);
2. activities (what was done);
3. people involvement (who was involved);
4. reactions (what they thought of it, immediate reactions as might be measured by exit surveys);
5. KASA change (i.e. changes after a period of time in the knowledge, attitudes, skills and aspirations of participants);
6. practice change (changes in the behavior of people); and
7. end results (or ultimate outcomes).

Bennett's model is not a program logic model; rather it is a conceptualization of the different stages of a project that evaluation should consider. Bennett's thesis was that instead of measuring the easy-to-measure things low down the hierarchy, evaluators should go up the hierarchy and consider all stages. Because measuring end results (ultimate outcomes or goals) may be difficult, especially in short timeframes, measuring KASA change and practice change provides interim indicators of the effectiveness (or likely effectiveness) of a project or policy. Together with a program logic model and a theory of change, some evidence of the likelihood of success is gained by having adequate people involvement (in terms of the target group), positive reactions from participants, some evidence of KASA change and some evidence of practice change. Empirical evidence may be hard to collect, especially if external conditions are changing, but stories of change from (a selection of) participants can readily be collected. If the majority of participants report that the activity has led to KASA change and has led to practice change, then that is reasonable evidence of success.

A BRIEF OVERVIEW OF QUALITATIVE METHODS USED IN EVALUATION

Qualitative evaluation uses a wide range of standard social research methods (i.e. ways of collecting data) and an ever-increasing number of innovative qualitative social research methods, including:

- Open-ended questions in structured questionnaires
- Semi-structured and in-depth interviews with key informants, undertaken in person, by phone and now increasingly by Skype
- Group interviews (essentially interviews with several people at the same time)
- Focus groups (groups discussions that are actively facilitated to focus on specific topics and where the discussion in the group is an important part of the process)

- Workshops, often with table-based or group exercises, including variants such as World Café (Brown and Isaacs, 2005)
- Role-plays and games
- Expert panels, citizen's juries (Smith and Wales, 2000) and other deliberative methods (Hartz-Karp and Pope 2011)
- Document analysis (i.e. analysing all available documents, photos, letters, emails, and other outputs of the activities associated with a project or program)
- Go-along interviews (Carpiano, 2009), *in situ* interviewing, shadowing (where the researcher/interviewer follows the participant as they go about their normal daily business)
- Story-telling with stories written down, audio-recorded or video-recorded
- Photo-elicitation (photovoice) and other visual 'triggering' techniques to stimulate participants to recall and articulate their thoughts and/or stories about certain topics (see Wang and Burris 1997)
- Mental models, mind maps and mud maps (i.e. a representational diagram showing the interconnections between related concept) – sometimes called spidergrams
- Participant diaries, logbooks and audio or visual recordings of reflections/comments as soon after they happen as practical
- Observation (sometimes aided by video recording)
- Participant observation and other experiential techniques
- Researcher diarizing (systematically recording notes in a field notebook).

There is a wide range of methodologies (ways of organizing data) and frameworks for thinking about information. There are also numerous theoretical frameworks (ways of interpreting data). In general, however, irrespective of the methodology or theoretical framework applied, the above techniques for collecting data tend to be utilized. In all social research, but especially in qualitative methods, partly as a quality-control mechanism, the use of multiple methods is common (also called 'mixed methods', especially when in conjunction with quantitative methods), and is methodologically preferred on the basis of a concept called 'triangulation' – that different methods should be used, with different sources of data and from different perspectives.

It should be noted that there are many approaches that combine theoretical understandings, methodologies and specific techniques into holistic and coherent frameworks that are suitable for particular situations. Many of the tools/methods used in these approaches can be very creative, which makes the construction of a comprehensive list of methods quite difficult. Some examples of these approaches that have an evaluative dimension include Rapid Rural Appraisal and Participatory Action Research (Chambers, 1994), Appreciative Inquiry (Cooperrider et al., 2008) and Outcomes Mapping (Earl et al., 2001).

THE ADDED VALUE OF QUALITATIVE METHODS IN EVALUATION

In general, qualitative research tends to:

- focus on meaning and on the 'why' rather than on 'how many'
- focus on issues where understanding is required rather than on confirming prior hypotheses
- be inductive rather than deductive
- be open rather than closed
- seek to discover the 'insider' rather than 'outsider' perspective
- be person-centered rather than variable-centered
- take a humanistic or phenomenological perspective rather than a positivist perspective
- be constructivist rather than naturalistic
- promote joint learning by all participants rather than just the learning of the researcher
- be holistic rather than particularistic
- be contextual (situated, embedded) rather than decontextual (distant, removed or detached). (Adapted primarily from Padgett, 2012 and a range of other sources)

In evaluation, qualitative research is used specifically to consider the why and how questions that quantitative methods typically can not answer, for example:

- Why does the program work (or not work)?
- How does the program achieve its goals?
- Why does it work for some clients (or in some situations) and not others?
- What are/were the needs of the clients that were not anticipated by program developers?
- What were the additional unintended and/or unexpected positive or negative consequences? (Adapted from Royce et al., 2001 and Anastas, 2004)

Thus qualitative methods are a valuable and important contribution to project and program evaluation, especially when the focus is formative (asking how the project or program can be improved) rather than summative (asking whether a project or program was a success).

A SHORT HISTORY OF QUALITATIVE EVALUATION

Guba and Lincoln (1989) describe four 'generations' in the history of evaluation: (1) measurement; (2) description; (3) judgment; and (4) constructivist evaluation. Perhaps a better word would have been paradigms or modalities rather than generations. A major criticism of this temporal overview of the field is that the extent to which the phases actually existed as discrete time periods is exaggerated – as can be demonstrated by the observation that the earlier generations are still very much in existence. Nevertheless, after a period of methodology battles or paradigm wars (Wadsworth, 2005), the dominant culture of evaluation is changing. While the division between summative evaluation and formative evaluation is widely accepted, evaluation is developing a greater focus as being creative (Patton, 1981), qualitative (Patton, 1990), participatory (Jackson and Kassam, 1998; Whitmore, 1998), utilization-focused (Patton, 1997),

constructivist or fourth generation (Guba and Lincoln, 1989), empowering or fifth generation (Fetterman et al., 1995; Laughlin and Broadbent, 1996; Fetterman, 2000), and as a form of action research (Whyte, 1990). Evaluation is now viewed as research for informing decision making at all phases of the project, program or policy (Vanclay et al., 2004). Instead of just being *ex-post* assessment or audit, evaluation is now understood to contribute to all stages of project or program development. Rather than being solely the domain of independent experts, evaluation is now widely seen as a participatory approach that empowers and builds capacity within institutions and among all program and project partners. Evaluation is now seen as a form of action research that informs project and program design (Vanclay et al., 2004). Ongoing evaluation and adaptive management are an essential part of being innovative and a learning organization. Evaluation is your friend.

Monitoring and evaluating the performance of rural development programs in developing countries has been an area where much innovation has taken place in evaluation methods, largely because of the inappropriateness of many traditional quantitative means of evaluation and because of the strong interest by funders in knowing whether their funding was being used effectively (Guijt et al., 2011). Many external evaluators and program managers have grappled with how to design fair methods that adequately capture the changes brought about by development interventions, especially when empirical indicators were not available, were too broad or not sufficiently sensitive to change.

While there were several story-based approaches to evaluation in the 1990s, two of the main proponents were Rick Davies and Jessica Dart. Although originally going by a variety of terms, in 2000 Davies and Dart settled on 'Most Significant Change Technique' as the term for the emergent method (Dart and Davies, 2003; Davies and Dart, 2005). Since then, the method has become firmly established in the evaluation and development cooperation professions, and as at September 2014 the term 'most significant change technique' had over 5000 hits on Google. Some evaluations using this approach include Wilder and Walpole (2008) and Waters et al. (2011).

Performance Story Reporting (PSR) is similar to and derives from the Most Significant Change Technique. Dart attributes the actual name 'performance story' to John Mayne of the Canadian Auditor General's Office (Mayne, 2004; Dart and Mayne, 2005). Arguably 'Performance Story Reporting' overcomes some of the criticisms of 'Most Significant Change Technique' (see Willetts and Crawford, 2007) and in particular removes concern about bias implied by the name and the (mis)perception that the method looked only for positive stories with atypical results. PSR has improved procedures to MSC, which means that it provides a more sophisticated, yet still workable, tool. While not yet widely represented in the scholarly literature, it is well known in the evaluation field (see Dart and Mayne, 2005) and as at September 2014 had over 3400 hits on Google.

Collaborative Outcomes Reporting (COR) is the latest incarnation of these story-based approaches to evaluation. Similar to PSR (in fact Dart provides her PSR work as examples of COR), the technique was renamed because of resistance to the word 'story' by some evaluation clients (Dart, pers. com., 20 April 2011) and because COR emphasizes integrating empirical and qualitative data and does not rely on the story alone. There is a range of other refinements, but these are minor, so information about

MSC and PSR is essentially still relevant to COR. As at September 2014, 'collaborative outcome reporting' had 1200 hits on Google (well up from only 50 or so in early 2012) and had been presented to a number of evaluation conferences.

There are various other story-based or narrative approaches to evaluation championed by many people, including Wadsworth et al. (2007), Kurtz (2009), Maxson et al. (2010) and Withers (2010). Stories and vignettes are also widely used in social research in a range of ways (see, e.g.,Dare et al., 2011; Howden and Vanclay, 2000; Hughes and Huby, 2002; Sandelowski, 1991; Vanclay and Enticott, 2011).

DESIGNING AND CONDUCTING A STORY-BASED APPROACH TO QUALITATIVE EVALUATION

Story-based evaluation approaches aim to be a rigorous qualitative method of reporting the impacts of projects and programs through stories. Intended to be used in conjunction with the reporting of specific empirical indicators where they are available, story-based approaches are especially useful to capture the broader social benefits of programs, particularly in situations where empirical attribution may be difficult. Mayne (2004, pp. 49–50) considers that:

> There are a variety of ways to present a performance story. All involve a mix of quantitative evidence that certain outputs and outcomes have occurred as well as narrative discussion and further evidence of the contributions made at various points along the results chain, all described within some context. A performance story sets out to convince a skeptical reader that the activities undertaken by the program have indeed made a difference – that the expectations chain has, at least to some extent, been realized, along with any significant unintended results.

Mayne (2004, p. 50) outlines the elements of a performance story as follows:

What is the context?
- the overall setting of the program (description, objectives, resources)
- the results chain (program theory)
- the risks faced

What was expected to be accomplished at what cost?
- statement of the (clear and concrete) outputs and outcomes expected
- planned spending

What was accomplished in light of these expectations?
- the relevant outputs delivered at what cost
- the outcomes realized related to the expectations
- a discussion of the evidence available demonstrating the contribution made by the program to those outcomes

What was learned and what will be done next?
- a discussion of what will be done differently as a result of what was achieved

What was done to assure quality data?

● a description of what the organization does to ensure the quality of the data and informa-
tion reported
The main story line of a performance story is how well the program has performed in rela-
tion to what was expected and what will now be done differently to better ensure future
performance.

The big difference between the approach by John Mayne and that taken by Jess Dart and
Rick Davies is where the stories come from and how they are constructed. For Mayne (at
the time in the Canadian Auditor General's Office), a performance story was an effective
way for the evaluator to provide information to a client, or how the Auditor General
might report to Parliament. The story technique was an effective means of expressing
complex information in a more credible and meaningful way to the target audience.

In contrast, MSC, PSR and COR all utilize the stories of participants themselves
(i.e. the program or project beneficiaries) in the report. These techniques have a
procedure for collecting stories and for selecting those that provide a good example of
the success of the project or program.

A REAL APPLICATION OF PERFORMANCE STORY REPORTING AT THE PROGRAM LEVEL

The Australian government funds a range of natural resource management (NRM)
programs and projects collectively known as the 'Caring for Our Country' program (see
http://www.nrm.gov.au/ and for a history see Hajkowicz, 2009). For the five-year period
from 2008 to 2013, the total value of the government's investment in this program is
intended to be about AU$2.25 billion (Commonwealth of Australia, 2008). Investments
in the two previous five-year periods totaled $1.5 billion and $1.3 billion respectively
(Auditor General, 2008). The program supports a range of disparate projects often in
conjunction with local community groups and may include small grant programs, co-
funding programs, and support for project staff and project costs.

While the objectives are now very clear (Commonwealth of Australia, 2008) and there
is a clearly elaborated program logic (Roughley, 2009), earlier versions of the program
did not have clearly identified intended outcomes. While anecdotal evidence suggests
that there was much benefit from the program (Curtis and Lockwood, 2000; Prager and
Vanclay, 2010), various official reports were dubious about the benefits. The Auditor
General's (2008, p. 16) report, for example, concluded that there were 'significant areas of
noncompliance by State agencies', and that 'the quality and measurability of the targets in
the regional plans is an issue for attention and . . . should be considered nationally – espe-
cially as the absence of sufficient scientific data has limited the ability of regional bodies
to link the targets in their plans to program outcomes'. The report went on:

14. There is evidence that activities are occurring 'on the ground'. For example, Environment's
2006–07 Annual Report commented that the programs have 'helped to protect over eight million
hectares of wetlands, have treated over 600 000 hectares of land to reduce salinity and erosion,
and have involved some 800 000 volunteers in on ground conservation work'. However at the
present time it is not possible to report meaningfully on the extent to which these outputs con-
tribute to the outcomes sought by government. There are long lead times for national outcomes
and delays in signing bilateral agreements did not help this process. The absence of consistently

validated data, the lack of agreement on performance indicators and any intermediate outcomes has significantly limited the quality of the reporting process.
15. Overall, the ANAO [Australian National Audit Office] considers the information reported in the DAFF [Department of Agriculture, Forestry and Fisheries] and NHT [Natural Heritage Trust] Annual Reports has been insufficient to make an informed judgement as to the progress of the programs towards either outcomes or intermediate outcomes. There is little evidence as yet that the programs are adequately achieving the anticipated national outcomes or giving sufficient attention to the 'radically altered and degraded Australian landscape' highlighted in the 1996 *Australia: State of the Environment* Report. Performance measurement has been an ongoing issue covered by three previous ANAO audits since 1996–97 and should be a priority for attention in the lead up to NHT3. (Auditor General, 2008, p. 16)

This context of official concern about the alleged benefits of the program, but strong public and political support for it, led to a real need to prove that the program was being successful, especially in the knowledge that empirical indicators were unlikely to reveal results in the short term.

Jessica Dart had been experimenting with the most significant change technique and performance story reporting for some time, first in the early to mid-1990s in developing-country contexts and later in an agricultural extension context in Australia. After working with the Department of Primary Industries in Victoria (Australia) for three years (and completing a PhD at Melbourne University in 2000), she established a consulting company, Clear Horizon. Jess's work was known by various people in government, especially in natural resource management circles, and it became evident that performance story reporting might be a good way to assist the Australian government in its need to capture the impacts of its investments in natural resource management (NRM).

The Australian government's Bureau of Rural Sciences (now part of the Australian Bureau of Agricultural and Resource Economics and Sciences, http://www.abares.gov.au) conducted a feasibility study of the use of story-based approaches by conducting four pilots (Carr and Woodhams, 2008). That 'independent review' considered three questions:

1) Are qualitative approaches such as MSC a) useful and b) appropriate as evidence of outcomes, including intermediate and other outcomes?
2) What are the strengths and weaknesses of using PSR to report by outcomes?
3) Could MSC and PSR be used by NRM regions in Australia as a form of participatory evaluation for producing program performance reports by outcomes? (Ibid., p. 3)

The report concluded:

Qualitative approaches to participatory evaluation such as MSC are both useful and appropriate as evidence of outcomes at multiple levels in NRM program logic hierarchies, including intermediate outcomes. Not only are qualitative approaches a valuable source of evidence of the changing human dimension of NRM, they are frequently a profound source of insight and sometimes the only kind of evidence available of the type of practice and attitudinal changes taking place.
. . .
Four key strengths of PSR were mentioned in reflective interviews: engagement, capacity building, problem-sharing and adaptive learning. These strengths were primarily associated with the MSC phase of the PSR process. Identifying and engaging evaluation stakeholders

was seen as a major strength of PSR and was the strength most frequently mentioned by interviewees. Many of the regional staff who took part in the MSC process appreciated the chance to build relationships with resource managers and develop their personal interviewing skills. The MSC process also increased communication about shared experiences and approaches to NRM problems that, in turn, led to an adaptive approach to natural resource management.

Arguably, there are two other key strengths of PSR. First, it integrates qualitative and quantitative evidence. Second, performance story reports rely upon participatory processes using program logic, which allows progressive collection and testing of evidence throughout the life of the investment program.

There are three key challenges for PSR: time and resources; data, results and interpretation; and complexity and preparedness. The biggest challenge across all stages of PSR was a perceived lack of time and resources to conduct the PSR process. Such comments came from all interviewees, consultants, regions and Australian Government representatives. At the regional level, interviewees were concerned that the goodwill and involvement from resource managers and regional staff would start to wane if the process was repeated each year without sufficient resources or local incentive. (Ibid., p. 61)

As a result of the positive feasibility assessment, the performance story reporting approach was rolled out across Australia with the Australian government publishing a 'User Guide', *Developing a Performance Story Report* (Roughley and Dart, 2009). While several project-level performance story reports were prepared, few of them were made available on the Internet. One example that is available is the performance story report for a biodiversity program in the Mount Lofty Ranges and Fleurieu Peninsula region of South Australia (Dart and O'Connor, 2008). Figure 25.2 is a copy of a page from that report showing how the stories are presented.

The project-level performance story reports can be aggregated into higher-level evaluations. One example is the assessment of natural resource management outcomes in the State of South Australia for the period 2001 to 2008, which drew on several performance story reports that were conducted in that state (see DWLBC, 2009). The performance story reports can be presented in a variety of formats, as conventional consultancy reports, as websites, as DVDs, as posters or brochures (see Figure 25.3), or as mixed media. Figure 25.4 shows a magazine/newsletter representation of a report with links to sound recordings of the participants' stories.

HOW TO UNDERTAKE A PERFORMANCE STORY REPORT EVALUATION

The Australian government's User Guide (Roughley and Dart, 2009) and various other instruction manuals (e.g. Dart et al., 2000; Davies and Dart, 2005; Silver et al., 2009) provide ample advice on how to conduct evaluations using performance story reporting.

A performance story report provides (Roughley and Dart 2009, p. 7):

- a view of progress towards outcomes at a glance
- insight into what's working and what's not and why
- a succinct account of program achievements

Section 4: Instances of significant change

The following vignettes were chosen by the participants of the summit workshop as representing the most significant changes occurring as a result of investment in the Recovery Program. These vignettes were chosen out of three sets of vignettes (22 in total).

What is a vignette?

Vignettes are used to elicit responses, interpretations and judgments about a particular set of circumstances or context within a research setting. Typically used in the qualitative social sciences, vignettes offer a method for simulating complex events, outcomes and/or problems and use these to explore people's perceptions, opinions, beliefs and attitudes. In this case, the vignettes were extracted directly from interview transcripts, which were captured with digital audio recording. While edits were kept to a minimum, some text was removed, this is indicated by three dots between sentences.

Vignette #1: Changing mind set

When we came here 10 years ago and found out how valuable the swamps are, and someone said you could be milking cows off it. ...Part of the reason that this property was bought was so that we could use the swamp for the cattle to be in, and now it has completely changed around to where you are not allowed to let, where we won't be letting any cattle in at all. ...I remember a long time ago, someone saying that this may be coming, and at the time I was thinking, oh well, we won't be doing that because we need our swamps for our cows, we can't shut them off. But that was our mind set...

And I don't think we contemplated shutting off the swamps when we were fully milking cows because that area was just too important. ... I think the fact that we have stopped milking cows makes it a whole lot easier too, because the milking cows needed to have every amount of grass...

Now I am ready to fence off swamp, but I have to do it slowly because Dad doesn't see all this land getting fenced off. ... [the project officer] came out and we got a bit of a line where we were going to run the fence. When I told Dad about it, he wanted to shift the line by quite a few metres further into the swamp so that he didn't loose so much land, but I think we have come to an agreement out of that, of how to do it, and what we should be doing. And usually with my father, he grizzles about things at the start, but once you have done it he has friends coming around and he is out showing them how great it is.

I think it is changing, I think there's more people that are into swamps... Maybe the more people do it the more acceptable it will be for other people to do it....It is a bit of an issue but it is more acceptable for us, I guess, and for others, I guess the more you hear about it the more you learn, the easier it is.

I guess I know now by seeing the map, the Emu-wrens are probably only a few kilometres away, and if we can certainly fence off our swamp and revegetate it, there is a chance that the Emu-wrens would move onto our property. And think the visual effect of seeing the swamps revegetated will certainly make the property look better, probably not quite as bare as it is. Everything that you hear around the place all the time, is how much everything is changing in our environment. Maybe this is just our way of trying to do our bit to help. It could be nicer for the kids, if the farm stayed in the family, they might see the rewards out of what we have done, maybe the Emu-wren will still be around then.

This vignette was considered significant by the participants at the summit workshop for the following reasons:
- It provides a real farmer perspective.
- It demonstrates how increased environmental knowledge and awareness can lead to attitudinal change and then to behavioral change.
- Promotion of change of practice needs to be continuous but also opportunistic. In this case the change in land use created an opportunity to encourage habitat protection (fencing).

24

Source: Dart and O'Connor (2008, p. 24).

Figure 25.2   Example of the presentation of a performance story

- an understanding of the links between investment and intended results, at intervals throughout a program
- a way for organisations to answer some of their more strategic evaluation questions on an as-needs basis or to use as a 'kick-start' process to revise the monitoring system
- information for future investment strategies
- a structure for an organisation's annual report.

Roughley and Dart (2009, p. 12) suggest that a typical report is between 10 and 30 pages long and comprises five parts:

Program context – background information about the program and the context in which it operates (how the program began, its location, objectives and key strategies, funding sources, structure and expected achievements), as well as an outline of the objectives and boundaries of the performance story evaluation and a summary of key outcomes and what has been learned.
Evaluation methodology – a brief overview of the process used in undertaking the evaluation.
Results – a listing of the most relevant and rigorous sources of evidence against the outcomes from the program logic hierarchy. This includes data as well as stories of change which are excerpts from interviews that best illustrate change that has occurred as a result of the program.
Findings and implications – a discussion framed by the evaluation questions that covers how the study has illustrated the program's impact (intended and unintended outcomes), the progress it has made towards its expected outcomes and how it has contributed to the long-term outcomes of NRM or a large NRM initiative. This part also includes recommendations for applying the findings to future phases of the program.
Index – a list of all the sources of evidence considered in the evaluation, including any additional references and the categories of interviewees and study participants.

Roughley and Dart (2009, p. 15, slightly modified) describe the seven steps to produce a report:

Scoping – inception/planning meetings are held to determine what will be evaluated, develop the program logic (if not already existing), set evaluation questions, and identify existing evidence and people to be interviewed.
Evidence gathering – an evidence trawl is conducted to identify existing data that will provide best evidence for expected outcomes. This is followed by the social inquiry process, where interviews are conducted with people who can provide additional information about program outcomes. Specific questions are asked and recorded to provide stories of significant changes that have occurred as a result of the program.
Integrated data analysis – quantitative and qualitative data is analysed to identify evidence corresponding to the outcomes in the program logic and integrated within the results chart.
Expert panel – people with relevant expertise assess the evidence of outcomes that has been gathered. They judge and make statements about the extent to which the evidence is adequate to assess the progress the program is making towards its stated outcomes. The panel may also identify further evidence that may be needed to make a conclusive statement about the achievement of program outcomes. Following the panel meeting, the evaluator integrates all of the analysed evidence and assesses the amount and quality of evidence available for each outcome in the program logic to inform a draft set of recommendations.

AN EVALUATION OF
Ribbons of Blue: 2008/2009

How Ribbons of Blue has contributed to 'pro-environmental' values, understanding and behaviour in school students; an evaluation of outcomes.

"You can look at a river and think it is just water and trees and stuff but now we look at the water and know that all these things are happening." Year 5 student

"It's changed everything: their caring, understanding, appreciating the importance of the environment, it's not just academic." Regional primary teacher

Ribbons of Blue:

- is an environmental education program which aims to motivate and enable current and future generations to respect and take action for healthy WA waterways
- has been going for 20 years and during 2008 and 2009 was delivered in nine regional catchment areas by 8.9 FTEs
- received $1,809,530 in funding in 2008/2009 from South West Catchments Council, the Departments of Environment and Conservation, Education, Water, Swan River Trust; WA Catchments Council, WA Plantation Resources and Shire of Manjimup.
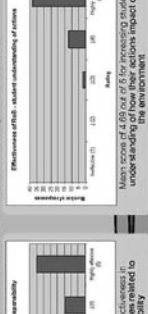
RIBBONS OF BLUE
REGIONS
2008/2009

By providing school educational activities through a range of curriculum areas, Ribbons of Blue aimed to enable students to:

- develop values related to environmental responsibility
- increase knowledge of local catchments and waterways
- increase understanding of how people impact on the health of waterways
- adopt behaviour/practices to improve the health of waterways.

This evaluation:

- was conducted by program staff, using the *Performance Story Reporting* process, with assistance from a consultant
- collected data from a survey of teachers (63 responses); three focus groups with students; 29 semi-structured interviews with teachers and coordinators' records, on 14 Key Evaluation Questions
- had an 'Expert Panel' review evidence and make outcome statements
- ran a 'Summit Workshop' of 22 stakeholders to discuss findings; agree on key outcomes and make recommendations.

Some limitations of this evaluation:

- Data was collected mainly from teachers (other key stakeholders not included).
- Teachers may want to 'protect' the program.
- Inexperience of program staff in conducting evaluation and interviews.

WHY ARE THE OUTCOMES THAT RIBBONS OF BLUE CONTRIBUTES TO IMPORTANT?

"I think that these kinds of outcomes are important if we are to produce environmentally aware citizens of the future who can push for and support environmentally friendly policies both in private enterprise and government institutions." Regional primary teacher

"A lot of the problems they hear about, like global warming and stuff, they can't really make a difference, and I try and impress on them that they don't have to solve all the world's problems, they can look after their little patch of it, that's a good start, so I think that's quite important too, that small changes can make a difference." Metropolitan secondary teacher

"It formulates who they are and where they are going and what they want their world to be like." Metropolitan primary teacher

"My classes feel ownership and have accepted responsibility for the Meelup Regional Park and have an ongoing love of the natural environment." Regional primary teacher

"If they don't take care of it then no-one else is going to." Metropolitan primary teacher

Contribution of Ribbons of Blue educational practice

"I believe the hands on learning, especially out in the field, is a far more effective way of learning. Seeing the excitement of kids looking at 'bugs' with binocular microscopes, and catching their own macros, is very rewarding for me as a teacher." Regional primary teacher

"Developing understanding and care for the world doesn't just come from a book – you need sight, smell, touch." Metropolitan primary teacher

"They need to know that it (science) is a real activity, its not just a subject you study at school, it has an impact on all sorts of aspects of their life. And if they don't get that understanding of science then we've failed fairly spectacularly as teachers." Metropolitan secondary teacher

CRITICAL SUCCESS FACTORS; regional coordinators and access to quality resources.

"Our programme focusing on the local environment through Ribbons of Blue has been possible because of the hands on activities with resourcing from the Teacher Resource Kit and the Regional Coordinator" Regional primary teacher

Key recommendations

Funding: explore all funding options to maintain/expand Ribbons of Blue in its current model of delivery; if no further funding is available, review and modify the Ribbons of Blue model of delivery while retaining as many of the qualities of the program that have been shown to be effective.

Delivery: develop new resources on specific topics (e.g. indigenous people and waterways); develop a means of 'capturing' and sharing how teachers have modified materials; improve marketing of products/services.

Impact: find a means of assessing the long-term impact of Ribbons of Blue on participants' actions/decisions later in life; develop 'tools' for directly measuring immediate changes in student understanding, skills, beliefs.

Ribbons of Blue — Our catchment: our future

Source: http://www.ribbonsofblue.wa.gov.au/component/option,com_docman/task,doc_download/gid,145/.

Figure 25.3   Example of the creative presentation of a performance story as a poster/brochure

Innovative Communication

NAILSMA Newsletter, October 2008

Please use this link to access this Newsletter online: http://www.nailsma.org.au/projects/oct08

NAILSMA Dugong and Marine Turtle Project News

Having trouble reading this newsletter? Click here to see it in your browser.

NAILSMA
North Australian Indigenous Land
& Sea Management Alliance

October 2008
In This Issue...

- Dugong and Marine Turtle Management Summit: Huge Success!
- Listen to what the participants had to say...
- Performance Story
- Nears Completion

Audio Interviews

- Bevan Bessen
- Brian Reid
- Charles David
- Daniel Oades
- Donna Kwon
- Eddie Sailor
- Frank Loban
- Ilse Kiessling
- Jane Blackwood
- Jenifer Rahmoy
- Joe Morrison
- Kathleen Mackie
- Kevin Scholes
- Lawrence Burke
- Leon Jackson
- Les Russell
- Libby Larsen
- Malaveçi Nona
- Mark Shadforth
- Moses Wabu
- Nathan Sandy
- Patrick White
- Paul Josif
- Pearson Wigness
- Phil Rist
- Richard Meister
- Robbie Salee
- Rod Kennett
- Roy Wigan
- Samuel Evans

Dugong and Marine Turtle Management Summit: Huge Success!

"Thumbs up!" say the participants of a recent Dugong and Marine Management Summit held by NAILSMA and hosted by the Kimberley Land Council and Bardi Jawi Rangers at Mudnunn on the Dampier Peninsular and brought together Indigenous Rangers and facilitators from as far away as Zenadh Kes (Torres Strait), Cape York, Gulf of Carpentaria and the Top End of the Northern Territory to a location just south of One Arm Point—a small community on the Western Australian coast that is home to the Bardi Jawi Rangers.

The Summit comes at the tail-end of the first phase of the Dugong and Marine Turtle Project and was an opportunity for Rangers and Communities involved in the Project to look back and share their successes over the past three years. It was also an opportunity to discuss issues and brainstorm new activities and aspirations that could be included in a future phase of the project.

Most significantly, all Project participants gave their overwhelming support to continue the Dugong and Marine Turtle Project with a focus on further increasing the capacity of Ranger Programs so that they can take their regional activity plans to the next level.

Also agreed to by the participants was an expansion of the

project to include more Indigenous Ranger units from across the North into the Project. The participants showed enthusiasm for the network to develop and see the opportunities for collaboration and sharing ideas across the regions as an important aspect of the Project.

A strong show of support for the Dugong and Marine Turtle Project by state and federal government was evident by the big mob of officials who participate in the four-day Summit. Heading the Department of the Environment, Water, Heritage and the Arts representation was Working On Country Assistant Secretary Kathleen Mackie who was accompanied by colleagues Donna Kwan, Ilse Kiessling, Jenifer Rahmoy and Les Russell. Representing the Great Barrier Marine Park Authority was Leon Jackson, the Australian Quarantine Inspection Service (NT) was Brian Reid, and Department of Indigenous Affairs (WA) was Jess Clements.

Listen to what the participants had to say about the Summit by clicking the pictures below.

Back to top

Listen to what the participants had to say...

Click the picture to be taken to the NAILSMA website where you can listen to audio from the participants recorded at the Summit.

The Summit ran from October 6 – 10

Project Partners and Participants

Kimberley Land Council
- Bardi-Jawi Rangers

Northern Land Council
- Dhimurru Rangers
- li-Anthawirriyarra Rangers

Carpentaria Land Council Aboriginal Corporation
- Wellesley Islands Rangers

Balkanu Cape York Development Corporation
- Injinoo Community
- Lockhart Community
- Pormpuraaw Rangers
- Hopevale Community

Torres Strait Regional Authority
- Mer (Murray) Islanders and Erub
- (Darnley) Islanders from Kemer Kemer Meriam Nation
- Badu Islanders
- Mabuiag Islanders and St. Pauls community from Mabuical Nation
- Moluk Kes (Boigu)

NAILSMA Dugong and Marine Turtle Project

- Email
  marine@cdu.edu.au
- Web
  www.nailsma.org.au

- Terrence Taylor
- Terry McCarthy
- Tonya Murray
- Vernon Yanner

Note: When originally created and viewed with an Internet connection, each speech bubble contained a link to an audio file which played the story of the person in the picture.

Source: http://www.savanna.org.au/nailsma/publications/downloads/Dugong-and-Marine-Turtle-Project-PSR.pdf.

Figure 25.4   Example of the creative presentation of a performance story as mixed media

Summit meeting – evaluation participants come together to consider and discuss the findings, nominate the stories that best illustrate the impact of the program and make recommendations for the program in future.

Integration, report and communications – the evaluator prepares the performance story report, which is a synthesis of all the above steps including recommendations from summit meeting participants. A plan is established to communicate the findings of the evaluation.

Revising the program logic – program managers, staff and other stakeholders meet to consider the report and revise the program logic as needed to plan for the next phase of the program. The next phase can incorporate the lessons and recommendations from the previous phase.

Figure 25.5 shows how the seven steps link to the difference parts of the report. The User Guide (Roughley and Dart, 2009) outlines each of the seven steps in some detail.

There are two critical steps in the process. The first is the process of recording stories (Step 2) and preparing them for presentation; and the second is the process of selecting stories to be included in the report (Step 4). A wide range of people should be interviewed. In addition to project participants and people involved at the project at all levels, strategic informants who can comment at a strategic level about how the program has been experienced or how it has addressed policy goals, such as funding agency representatives and policy officers from local, state or federal government, can also be interviewed. The User Guide provides guidelines for how the stories should be prepared.

The expert panel comprises a group of people charged with having oversight for ensuring that the body of evidence collected, including the stories of change, are credible and realistic examples of the changes that have been made.

FRANK COMMENTARY ON THE PERFORMANCE STORY REPORTING METHOD

Ironically, in comparison with the Most Significant Change Technique, in Performance Story Reporting the role of stories has been downplayed, and the importance of program logic and the capacity of PSR to be an integrative approach to evaluation using qualitative and quantitative data has been emphasized. This is even more so in Collaborative Outcome Reporting, the next incarnation of the method. It is likely that there was some skepticism about a story-based approach in official and scientific circles in Australia, and it is possible that, as a result, the method was (re)designed in an attempt to maximize its perceived credibility to the key stakeholders (i.e. the Auditor General and political detractors). The inclusion of an expert panel is one example of the attempt to increase the legitimacy of the method in the eyes of the cynics. For people committed to participatory approaches and/or who are accepting of qualitative methods, the expert panel is unnecessary and perhaps undesirable, and is not likely to be effective or useful. However, where an overseeing or monitoring function is needed, they could be used.

Curiously, PSR was used in Australia because there was no pre-existent program logic and little capacity to utilize data other than the performance stories. The unique value of the method that led to its use was its ability to be a retrospective assessment and to

DEVELOPING A PERFORMANCE STORY REPORT

Program context

Evaluation methodology

Results

Findings and implications

Index

STEP 1
Scoping
Community participants
Program leaders and managers
Evaluator

STEP 2
Evidence gathering
Evaluator
Community participants
Program leaders and managers

STEP 3
Integrated data analysis
Scientists across disciplines
Program leaders and managers
Community participants

STEP 4
Expert panel
Evaluator

STEP 5
Summit meeting
Evaluator
All program participants

STEP 6
Integration, report and Communications
Evaluator

STEP 7
Revising the Program logic
Program team

Source: Roughley and Dart (2009, p.13).

Figure 25.5   Steps in producing a performance story report

provide some evidence of change in the absence of baseline data. The stories of participants gave a sense of the multifactorial nature of the outcomes, were able to adjust to the confounding effects of external events (such as the drought most of Australia experienced for much of the decade), and detailed the unanticipated benefits that were experienced.

While the use of the method was arguably successful in the Australian context, there were some issues that could be addressed should the methodology be used again in the future. It also has some inherent limitations. There is potential for considerable variation in the way the stories are compiled, especially when multiple interviewers are used. It is necessary to ensure good training, supervision and monitoring of the interviewers so that they are relatively consistent in the way stories are collected and to ensure that the stories are effective as stories and valid as a fair reporting of participant experience. Given that the procedure allows for a degree of editing of the transcripts (see p. 37 of the User Guide for precise statements about this), it is quite likely that there will be variation in the extent to which this occurs.

From a social research perspective, the analysis should be done on all the available data using all interview transcripts that are available, not just the stories selected for inclusion in the report. Using qualitative data analysis software (e.g. NVivo) to code the original interview transcripts (rather than the enhanced stories) would be appropriate. The selected stories are intended only to exemplify the information in the evaluation report and do not constitute the data or the analysis. Therefore it might be argued that there is no harm in the stories being modified or enhanced, or even combined to produce indicative narratives (see Dare et al., 2011), rather than attempt to be faithful to the notion of being interview transcripts. The method of selecting and ratifying the stories used as vignettes in the Summit Meeting provides a quality control to ensure that the vignettes are legitimate and authentic even if they are not the actual words of a single person. The PSC method (as strictly outlined in the User Guide) is a hybrid approach and is confused in its methodological positioning. Conceivably the expert panel and the rules about the stories are about ensuring the external legitimacy of the process, rather than necessarily being about the integrity of the data or the analysis.

Notwithstanding the above criticisms, the key point is that story-based evaluation is an effective way of collecting evidence of change, especially in contexts where there are not adequate empirical indicators, where causality is hard to establish, and where there may be external factors that influence the outcomes. In these contexts, the stories of participants provide a means of determining success. Similar to all qualitative social research, the robustness of the method is established through the professionality of the researchers, and the consistency of stories from multiple sources (triangulation), allowing of course for different perspectives from different stakeholders. It is clear that story-based evaluation could be an appropriate approach in the context of EU cohesion funds, especially as an augmentation to other forms of evaluation.

ANSWERS TO SOME FREQUENTLY ASKED QUESTIONS

1.   Are there any ethical considerations that are particular to the approach?

There are no particular ethical considerations applying to story-based approaches to

evaluation than would also apply to other evaluation techniques or social research. The standard ethical principles that apply to all social research equally apply. It would be generally expected that:

- participants have given informed consent
- their participation is voluntary and that there has been no coercion or threatened or implied retribution for non-participation
- participants can withdraw at any time, and have their data removed from the analysis (where this is possible)
- there is full disclosure of funding sources
- there will be no harm to participants; that the researchers have given consideration about the potential for inadvertent harm; and where emotional distress and so on arise as a result of participation in the research, that the researcher is able to assist the participant to resolve this or to seek appropriate professional care
- interviewers and other staff demonstrate respect towards participants
- anonymity can be presumed, or, if it is intended to attribute statements, that the expressed permission for the use of names be provided, including where a person's identity is evident from the context (e.g. the mayor, governor, president of a particular organisation etc.)
- there be confidentiality with respect to all private matters (or when any such an undertaking is given)
- there be full disclosure of research methods used to enable replication of the research by another researcher, and to enable peer review of the adequacy and ethicality of the methodology, and to encourage critical self-reflections on the limitations of the methodology and the implications of this for the results and conclusions.

2.   Is it subject to bias or open to distortion or manipulation?

All forms of social research are affected by 'bias'. Bias is a technical term in social research meaning a systematic tendency to favor one outcome over another as an inherent feature of the methods used. It is different from random error. Bias can occur in terms of response bias to surveys (to what extent are people who answer surveys different to those who don't respond?). Bias can occur in situations where social desirability encourages people to understate or sometime overstate their experience (e.g. their weekly alcohol consumption). People's beliefs about what the research will be used for will affect their likely answers. Thus all forms of social research are affected by bias, and qualitative methods are also so affected.

It is highly likely, however, that qualitative methods are less subject to bias than many quantitative methods. First, it is part of the code of practice of qualitative researchers that they be acutely aware of how their techniques may cause bias and that they reflexively take steps to minimise bias. Second, in an interview where an interviewer has an inkling that the statements of the respondent are exaggerated, this can (and should) be annotated on the transcript.

In terms of the Performance Story Reporting method used in Australia, bias was minimized by the use of expert panels whose task was to ensure that the stories selected accu-

rately and fairly represented the typical experiences of the participants in the program activities.

3. *I can see how it works at the project/activity level; how does it apply at the program level?*

There are two responses to the question. The first is that stories can be collected from all participants, not only community participants in program activities, but also program managers and coordinators. The stories that are collectively assessed for the evaluation could come, for example, from the people in charge of the program in their own region.

The second response is that program evaluation can be seen as a type of meta-evaluation. Programs ultimately come down to activities on the ground that are coordinated at various levels. The way the evaluation of the program in Australia worked was that story-based evaluations were done at the regional level, with a report written for each region. The overarching program evaluation was then a meta-analysis of the various regional level Performance Story Reports.

4. *Does it work for summative evaluation as well as for formative evaluation, and can it measure 'impact'?*

Summative and formative are terms that are used to describe the purpose to which evaluation is being put. They represent a continuum rather than completely separate entities. Formative is focused on contributing to the ongoing development and improvement of the program, whereas summative attempts to be an *ex-post*, independent, objective assessment of outcomes that assesses whether the extent to which the program was responsible for (i.e. caused) the outcomes, and potentially undertaking a cost–benefit or return-on-investment calculation. For these reasons, summative evaluation tends to require empirical indicators, while formative evaluation tends to be qualitative.

Story-based approaches to evaluation, however, tend to be for summative evaluation purposes. Formative evaluation uses a wide range of qualitative methods and, because of the developmental process of the program, there usually is no issue that questions the validity of the qualitative data, especially in terms of views about how the program could be improved. Formalized story-based approaches tend to be used for summative evaluation in situations where there are no empirical indicators, no baseline measurements, no previously identified program logic, and complex multifactor and/or changing environments, and/or situations with long lag times between program/project activities and likely outcomes. In situations with long lag times, even if empirical indicators are available, they may not show movement in the political timescales necessary to make decisions about funding. In these situations, just as occurred in Australia with the need to evaluate the 'Caring for Our Country' program, performance story reporting provided a solution.

To repeat: story-based evaluation (and performance story reporting in particular) are forms of qualitative evaluation specifically adapted to be appropriate in the context of summative evaluation. As qualitative methods, they do not intend to provide empirical measures; however, as Dart emphasizes, performance story reporting and collaborative outcomes reporting are meant to be integrative approaches. Furthermore, the qualitative

research may generate ideas that could lead to the identification of possible empirical indicators.

CONCLUSION

There is ample testament to the power of stories. Tell a person an isolated alleged 'fact', and they wonder about its veracity. Tell a person a proper story, and it will likely be accepted. Stories are more engaging, more meaningful, more real; they convey information more effectively and are more likely to be remembered than facts.

An effective story must be a proper story. It can not be an inchoate amalgam of odd ideas. To be an effective story, it needs to conform to the standard basic elements of all stories. It needs to have a beginning, a middle and an ending. It needs to have a coherent and credible storyline running through it. It needs to be multidimensional, but the different components need to be connected and the causal relations between the components need to become clear in the course of the story. It needs to be personal and emotional. Kurtz (2009) provides much advice on how to construct good stories.

Telling stories as a means of effecting behavior change is an ancient art. Biblical parables, children's fables, classic mythology and good literature all seek to influence their readers. Using stories to understand, analyze and make sense of things is relatively recent, but has been part of strategic planning in business for some decades. Using stories as an evaluation methodology is even more recent, but has much appeal and, as demonstrated by the Australian experience, can be effectively implemented.

It is not intended that story-based approaches replace quantitative indicators where they are available. The intention is that the stories complement the quantitative indicators and that they add value to those indicators by providing meaning and interpretation. Big programs are subject to long lag times, and can be subject to the influence of multiple external influences. It is naïve to think that a list of simple indicators will reveal the complex processes taking place. Stories are therefore a much more effective way of understanding what is happening.

NOTE

REFERENCES

Anastas, J. (2004) Quality in qualitative evaluation: issues and possible answers, *Research on Social Work Practice*, **14**(1), 57–65.
Auditor General (2008) Regional Delivery Model for the Natural Heritage Trust and the National Action Plan for Salinity and Water Quality (Audit Report No.21 2007-08). Australian National Audit Office,

Canberra. Last accessed 11 April 2012, at http://www.anao.gov.au/uploads/documents/2007-08_Audit_Report_21.pdf.

Bennett, C. (1975) Up the hierarchy, *Journal of Extension*, **13**(2), 7–12.

Brown, J. and Isaacs, D. (2005) *The World Café: Shaping Our Futures Through Conversations that Matter*, San Francisco, CA: Berrett-Koehler Publishers.

Carpiano, R. (2009) Come take a walk with me: the 'go-along' interview as a novel method for studying the implications of place for health and well-being', *Health & Place*, **15**, 263–72.

Carr, A. and Woodhams, F. (2008) Monitoring, Evaluating, Reporting and Improving (MERI) Natural Resource Management: Making MERI in Australia, Canberra: Bureau of Rural Sciences.

Cash, D., Clark, W., Alcock, F., Dickson, N., Eckley, N., Guston, D., Jäger, J. and Mitchell, R. (2003) Knowledge systems for sustainable development, *Proceedings of the National Academy of Sciences USA*, **100**(14), 8086–91.

Chambers, R. (1994) The origins and practice of participatory rural appraisal, *World Development*, **22**(7), 953–69.

Commonwealth of Australia (2008) Caring for Our Country Outcomes 2008–2013. Canberra: Commonwealth of Australia,. Last accessed 11 April 2012, at http://www.nrm.gov.au/resources/index.html.

Cooperrider, D.L., Whitney, D. and Stavros, J.M. (2008) *Appreciative Inquiry Handbook* (2nd edn), Brunswick, OH: Crown Custom.

Curtis, A. and Lockwood, M. (2000) Landcare and catchment management in Australia: lessons for state-sponsored community participation, *Society & Natural Resources*, **13**(1), 61–73.

Dare, M., Vanclay, F. and Schirmer, J. (2011) Understanding community engagement in plantation forest management: insights from practitioner and community narratives, *Journal of Environmental Planning and Management*, **54**(9), 1149–68.

Dart, J. and Davies, R. (2003) A dialogical, story-based evaluation tool: the Most Significant Change Technique, *American Journal of Evaluation*, **24**(2), 137–55.

Dart, J. and Mayne, J. (2005) Performance story, in S. Mathison (ed.), *Encyclopedia of Evaluation*. Thousand Oaks, CA: Sage, pp. 307–8.

Dart, J. and O'Connor, P. (2008) Performance Story Report: A Study of the Mount Lofty Ranges Southern Emu-Wren and Fleurieu Peninsula Swamps Recovery Program and how it contributed to biodiversity outcomes in the Adelaide and Mount Lofty Ranges Natural Resources Management Region. Commonwealth of Australia, Canberra. Last accessed 11 April 2012, at http://nrmonline.nrm.gov.au/catalog/mql:2167.

Dart, J., Drysdale, G., Cole, D. and Saddington, M. (2000) The Most Significant Change approach for monitoring an Australian extension project, *PLA Notes*, **38**, 47–53 (International Institute for Environment and Development, London).

Davies, R. and Dart, J. (2005) *The 'Most Significant Change' (MSC) Technique: A Guide to its Use* (self-published manual for a consortium of development agencies). Last accessed 11 April 2012, at http://www.mande.co.uk/docs/MSCGuide.pdf.

Denning, S. (2007) *The Secret Language of Leadership: How Leaders inspire Action through Narrative*. San Francisco, CA: Jossey-Bass.

DWLBC (2009) *Our Changing Environment: Outcomes and Lessons from Investment in South Australia's Natural Resources 2001–2008*, Government of South Australia, through the Department of Water, Land and Biodiversity Conservation, Adelaide. Last accessed 11 April 2012, at http://www.nrm.sa.gov.au/Portals/0/NRMinSA/Our_changing_environment_full.pdf.

Earl, S., Carden, F. and Smutlyo, T. (2008) Outcome mapping: building learning and reflection into development programs. Ottawa: International Development Research Centre.

Esteves, A.M. and Vanclay, F. (2009) Social Development Needs Analysis as a tool for SIA to guide corporate–community investment: applications in the minerals industry, *Environmental Impact Assessment Review*, **29**(2), 137–45.

Esteves, A.M., Franks, D. and Vanclay, F. (2012) Social impact assessment: the state of the art, *Impact Assessment and Project Appraisal*, **30**(1), 35–44.

EVALSED (2009) Introduction to CIE, EVALSED: the resource for the evaluation of socioeconomic development. Last accessed 11 April 2012, at http://ec.europa.eu/regional_policy/sources/docgener/evaluation/evalsed/sourcebooks/method_techniques/counterfactual_impact_evaluation/introduction_to_cie/index_en.htm.

Feller, I. (2007) Mapping the frontiers of evaluation of public sector R&D programs, *Science and Public Policy*, **34**(10), 681–90.

Fettermam, D.M. (2000) *Foundations of Empowerment Evaluation*. Thousand Oaks, CA: Sage.

Fetterman, D.M., Kaftarian, S.J. and Wandersman, A. (eds) (1995) *Empowerment Evaluation*. Thousand Oaks, CA: Sage.

Fisher, W.R. (1989) *Human Communication as Narration: Toward a Philosophy of Reason, Value, and Action*. Columbia, SC: University of South Carolina Press.

Greene, J.C. (2000) Qualitative program evaluation: practice and promise, in NK. Denzin and Y.S. Lincoln (eds), *Handbook of Qualitative Research*. Thousand Oaks, CA: Sage, pp. 530–44.

Guba, E. and Lincoln, Y. (1989) *Fourth Generation Evaluation*. Newbury Park, CA: Sage.

Guijt, I., Brouwers, J., Kusters, C., Prins, E. and Zeynalova, B. (2011) *Evaluation Revisited: Improving the Quality of Evaluative Practice by Embracing Complexity* (Conference Report). Centre for Development Innovation, Wageningen University & Research Centre. Last accessed 11 April 2012, at http://capacity.org/capacity/export/sites/capacity/documents/topic-readings/110412-evaluation-revisited-may-2010_small-version.pdf.

Hajkowicz, S. (2009) The evolution of Australia's natural resource management programs: towards improved targeting and evaluation of investments, *Land Use Policy*, **26**(2), 471–8.

Hartz-Karp, J. and Pope, J. (2011) Enhancing effectiveness through deliberative democracy, in F. Vanclay and A.M. Esteves (eds), *New Directions in Social Impact Assessment: Conceptual and Methodological Advances*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 253–72.

Howden, P. and Vanclay, F. (2000) Mythologization of farming styles in Australian broadacre cropping, *Rural Sociology*, **65**(2), 295–310.

Hughes, R. and Huby, M. (2002) The application of vignettes in social and nursing research, *Journal of Advanced Nursing*, **37**(4), 382–6.

Jackson, E.T. and Kassam, Y. (eds) (1998) *Knowledge Shared: Participatory Evaluation in Development Cooperation*. West Hartford, CT: Kumarian Press.

João, E., Vanclay, F. and den Broder, L. (2011) Emphasising enhancement in all forms of impact assessment, *Impact Assessment & Project Appraisal*, **29**(3), 170–80.

Kurtz, C. (2009) Working with Stories in your Community or Organization. Last accessed 11 April 2012, at http://www.workingwithstories.org.

Laughlin, R. and Broadbent, J. (1996) Redesigning fourth generation evaluation, *Evaluation*, **2**(4), 431–51.

Leeuw, F. and Vaessen, J. (2009) Impact Evaluations and Development: NONIE Guidance on Impact Evaluation. The Network of Networks on Impact Evaluation (c/o Independent Evaluation Group, The World Bank), Washington DC. Last accessed 11 April 2012, at http://siteresources.worldbank.org/EXTOED/Resources/nonie_guidance.pdf.

Mayne, J. (1999) Addressing attribution through contribution analysis: using performance measures sensibly (Discussion Paper), Office of Auditor-General of Canada. Last accessed 11 April 2012, at http://dsp-psd.pwgsc.gc.ca/Collection/FA3-31-1999E.pdf.

Mayne, J. (2004) Reporting on outcomes: setting performance expectations and telling performance stories, *The Canadian Journal of Program Evaluation*, **19**(1), 31–60.

Maxson, M., Guijt, I. et al. (2010) The 'Real Book' for story evaluation methods, GlobalGiving Foundation. Last accessed 11 April 2012, at http://www.globalgiving.org/jcr-content/gg/landing-pages/story-tools/files/-story-real-book--2010.pdf.

Padgett, D.K. (2012) *Qualitative and Mixed Methods in Public Health*, Thousand Oaks, CA: Sage.

Patton, M.Q. (1981) *Creative Evaluation*. Beverly Hills, CA: Sage.

Patton, M.Q. (1990) *Qualitative Evaluation and Research Methods*. Newbury Park, CA: Sage.

Patton, M.Q. (1997) *Utilization-Focused Evaluation*. Thousand Oaks, CA: Sage.

Prager, K. and Vanclay, F. (2010) Landcare in Australia and Germany: comparing structures and policies for community engagement in natural resource management, *Ecological Management & Restoration*, **11**(3), 187–93.

Roese, N. (1997) Counterfactual thinking, *Psychological Bulletin*, **121**(1), 133–48.

Roughley, A. (2009) *Developing and Using Program Logic in Natural Resource Management: User Guide*. Australian Government, Canberra. Last accessed 11 April 2012, at http://nrmonline.nrm.gov.au/catalog/mql:2164.

Roughley, A. and Dart, J. (2009) *Developing a Performance Story Report User Guide*. Australian Government, Canberra. Last accessed 11April 2012, at http://nrmonline.nrm.gov.au/catalog/mql:2162.

Royce, D., Thyer, B.T., Padgett, D.K. and Logan, T.K. (2001) *Program Evaluation: An Introduction* (3rd edn). Pacific Grove, CA: Brooks/Cole.

Sandelowski, M. (1991) Telling stories: narrative approaches in qualitative research, *Image: The Journal of Nursing Scholarship*, **23**(3), 161–6.

Shaw, G., Brown, R. and Bromiley, P. (1998) Strategic stories: how 3M is rewriting business planning, *Harvard Business Review*, **76**(3), 41–50.

Silver, M., Richards, R. and Cowell, S. (2009) *A Guide to Designing and Implementing Performance Reporting to increase the Confidence of Conservation Investors*. Bush Heritage Australia, Melbourne. Last accessed 11 April 2012, at http://lwa.gov.au/files/products/reporting/pn30209/pn30209.pdf.

Smith, G. and Wales, C. (2000) Citizens' juries and deliberative democracy, *Political Studies*, **48**, 51–65.

Vanclay, F. (2003) International principles for social impact assessment, *Impact Assessment & Project Appraisal*, **21**(1), 5–11.

Vanclay, F. (2012) The potential application of Social Impact Assessment in integrated coastal zone management, *Ocean & Coastal Management*, **68**, 149–56.

Vanclay, F. (in press), The potential application of qualitative evaluation methods in European regional development: reflections on the use of Performance Story Reporting in Australian natural resource management, *Regional Studies*, http://dx.doi.org/10.1080/00343404.2013.837998.

Vanclay, F. and Enticott, G. (2011) The role and functioning of cultural scripts in farming and agriculture, *Sociologia Ruralis*, **51**(3), 256–71.

Vanclay, F., Lane, R., Wills, J., Coates, I. and Lucas, D. (2004) 'Committing to place' and evaluating the higher purpose: increasing engagement in natural resource management through museum outreach and educational activities, *Journal of Environmental Assessment Policy & Management*, **6**(4), 539–64.

Wadsworth, Y. (2005) 'Gouldner's child?': some reflections on sociology and participatory action research, *Journal of Sociology*, **41**(3), 267–84.

Wadsworth, Y., Wierenga, A. and Wilson, G. (2007) *Writing Narrative Action Evaluation Reports in Health Promotion: Manual of Guidelines, Resources, Case Studies and QuickGuide* (2nd edn), State of Victoria, Department of Human Services and the University of Melbourne, Australia.

Wang, C. and Burris, M. (1997) Photovoice: concept, methodology, and use for participatory needs assessment, *Health Education & Behavior*, **24**(3), 369–87.

Waters, D., James, R. and Darby, J. (2011) Health promoting community radio in rural Bali: an impact evaluation, *Rural and Remote Health*, **11**(1), article 1555 (online journal). Last accessed 11 April 2012, at http://www.rrh.org.au/articles/subviewnew.asp?ArticleID=1555.

Whitmore, E. (ed.) (1998) *Understanding and Practicing Participatory Evaluation: New Directions for Evaluation*. San Francisco, CA: Jossey-Bass.

Whyte, W.F. (ed.) (1990) *Participatory Action Research*. Newbury Park, CA: Sage.

Wilder, L. and Walpole, M. (2008) Measuring social impacts in conservation: experience of using the Most Significant Change method, *Oryx*, **42**(4), 529–38.

Willetts, J. and Crawford, P. (2007) The most significant lessons about the Most Significant Change Technique, *Development in Practice*, **17**(3), 367–79.

Withers, D. (2010) A narrative evaluation of the City of Surrey's Cultural Capital of Canada Award Program. Last accessed 11 April 2012, at https://files.me.com/denisewithers/73mf37.