University of Groningen

The Potential Application of Qualitative Evaluation Methods in European Regional Development

Vanclay, Frank

Published in:
Regional Studies

DOI:
[10.1080/00343404.2013.837998](https://doi.org/10.1080/00343404.2013.837998)

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

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Vanclay, F. (2015). 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*, 49(8), 1326-1339. <https://doi.org/10.1080/00343404.2013.837998>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

The Potential Application of Qualitative Evaluation Methods in European Regional Development: Reflections on the Use of Performance Story Reporting in Australian Natural Resource Management

FRANK VANCLAY

*Department of Cultural Geography, Faculty of Spatial Sciences, University of Groningen, PO Box 800,
NL-9700AV Groningen, the Netherlands. Email: frank.vanclay@rug.nl*

(Received May 2012; in revised form August 2013)

VANCLAY F. 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*. This paper argues that qualitative evaluation methods potentially have a useful role in the assessment of regional development projects in Europe. It outlines several evaluation methods used in outcomes assessment, specifically Most Significant Change Technique, Performance Story Reporting and Collaborative Outcomes Reporting. An example of the practical application of these methods in Australia is provided along with a consideration of their applicability in the European context. The paper discusses issues related to the evaluation of rural and regional development programmes and concludes that qualitative evaluation using story-based approaches provides a rigorous way of assessing the performance of projects and programmes.

Qualitative evaluation Story-based approaches Narrative evaluation Outcomes assessment Retrospective evaluation Impact evaluation

VANCLAY F. 质化评估方法在欧洲区域发展上的可能应用：奥地利自然资源管理使用绩效叙述报告的反思，*区域研究*。本文主张，质性研究评估方法在评价欧洲区域发展计画上具有潜在的用处。本文概述运用于成果评量的部分评估方法，特别是“最重要的变化技术”、“绩效叙述报告”以及“协作成果报告”。我将提供这些方法在奥地利的实际运用案例，并考量它们在欧洲脉络的适用性。本文探讨乡村与区域发展计画评估的相关议题，并做结道：运用根据叙述方法的质化评估，提供了评价计画与方案绩效的严谨方法。

质化评估 根据叙述的方法 叙事评估 成果评量 回溯性评估 冲击评估

VANCLAY F. L'application éventuelle des méthodes d'évaluation qualitative dans l'aménagement du territoire en Europe: des réflexions sur l'emploi du Performance Story Reporting quant à la gestion des ressources naturelles en Australie, *Regional Studies*. Cet article affirme que les méthodes d'évaluation qualitative jouent un rôle utile dans l'évaluation des projets d'aménagement du territoire en Europe. On esquisse plusieurs méthodes d'évaluation employées dans l'évaluation des résultats, notamment la Most Significant Change Technique (Technique de changement le plus significatif), Performance Story Reporting (Tableau de bord permettant de mesurer la progression vers les objectifs et les priorités) et Collaborative Outcomes Reporting (Rapport collaboratif sur les résultats). On fournit un exemple de l'application pratique de ces méthodes en Australie et on prend également en considération leur applicabilité dans un cadre européen. L'article discute des questions qui se rapportent à l'évaluation des programmes d'aménagement ruralo-régional et conclut qu'une évaluation qualitative qui emploie des façons axées sur les histoires fournit un moyen exhaustif d'évaluer les projets et les programmes.

Évaluation Qualitative Façons axées sur les histoires Évaluation narrative Évaluation des résultats Évaluation rétrospective Évaluation d'impact

VANCLAY F. Die potenzielle Anwendung von qualitativen Bewertungsmethoden in der europäischen Regionalentwicklung: Überlegungen über den Einsatz von Performance Story Reporting im Bereich der Naturressourcenverwaltung in Australien, *Regional Studies*. In diesem Beitrag wird die These aufgestellt, dass qualitative Bewertungsmethoden eine potenziell nützliche Rolle bei der Bewertung von regionalen Entwicklungsprojekten in Europa spielen können. Beschrieben werden mehrere Methoden zur Ergebnisbewertung: die 'Most Significant Change Technique', das 'Performance Story Reporting' und das

‘Collaborative Outcomes Reporting’. Der Beitrag enthält ein Beispiel für die praktische Anwendung dieser Methoden in Australien sowie eine Untersuchung ihrer Anwendbarkeit auf den europäischen Kontext. Ebenso erörtert werden Aspekte im Zusammenhang mit der Bewertung von Programmen zur ländlichen und regionalen Entwicklung; das Fazit lautet, dass eine qualitative Bewertung mit Hilfe von Story-Ansätzen eine präzise Methode zur Untersuchung der Leistung von Projekten und Programmen bietet.

Qualitative Bewertung Story-Ansätze Narrative Bewertung Ergebnisbewertung Retrospektive Bewertung
Bewertung von Auswirkungen

VANCLAY F. La posible aplicación de métodos cualitativos de evaluación en el desarrollo regional europeo: reflexiones sobre el uso de los informes sobre el desempeño en la gestión de recursos naturales en Australia, *Regional Studies*. En este artículo se argumenta que los métodos de evaluación cualitativa tienen potencialmente un papel útil en la valoración de los proyectos de desarrollo regional en Europa. Se destacan varios métodos de evaluación que se utilizan en la valoración de resultados, especialmente la ‘técnica del cambio más significativo’, los ‘informes sobre el desempeño’ y los ‘informes sobre los resultados de colaboración’. Se muestra un ejemplo de la aplicación práctica de estos métodos en Australia junto con una consideración de su aplicabilidad en un contexto europeo. Se analizan las cuestiones relacionadas con la evaluación de los programas de desarrollo rural y regional y se concluye que la evaluación cualitativa, en la que se utilizan planteamientos basados en historias, ofrece un método riguroso de valorar el desempeño de proyectos y programas.

Evaluación cualitativa Planteamientos basados en historias Evaluación narrativa Valoración de resultados Evaluación retrospectiva Evaluación de repercusiones

JEL classifications: B, B0, B00, B5, B50, H, H0, H00, Q, Q0, Q01

INTRODUCTION: THE NEED FOR QUALITATIVE EVALUATION

The attempt to identify what works and why are perennial questions for evaluators, programme and project managers, funding agencies, and policy-makers (GREENE, 2000; FELLER, 2007). Policies, programmes, plans and projects (hereafter all ‘programmes’ 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 programmes 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 programme, to determine whether subsequent tranche payments should be made, and also to assist in decision-making about whether similar programmes should be funded.

Evaluation in such circumstances is complex. How can the achievement of goals be assessed if they are long-term? Evaluation cannot wait years to determine whether a programme has been successful. Thus, evaluation needs to consider carefully the programme 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 programmes, especially long-term ones, should incorporate a degree of adaptive management or reflexivity into them 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 programme has effective monitoring and is capable of adapting to feedback as it progresses. Depending on the context, it may also be the case that various external factors have changed and the

original goals and/or programme logic need to be revised to accommodate the changed circumstances. Any programme seeking to contribute to high-level goals like enhanced community well-being, social sustainability, regional development potential, innovativeness, etc., is likely to be affected by a changing context. Therefore, a key factor for success (and thus for evaluation) is the ability of the programme to be responsive to change. How is this adaptivity to be evaluated?

It is increasingly evident that standard ex-post quantitative evaluation techniques are not adequate to deal with these matters. Standard approaches to evaluation cannot deal with long lag times, they cannot cope with multi-causality, and they cannot cope with a changing operating environment. Qualitative methods, on the other hand, offer a way of collecting evidence about the performance of a programme or project. They also enable the collection of feedback to assist in modifying the programme. They can work in tandem with quantitative indicators, playing a complementary but different although equally important role. This paper, which is based on a report developed at the request of the European Commission’s Directorate General for Regional Policy (VANCLAY, 2012a), explores issues associated with the evaluation of regional development programmes in Europe and elsewhere. Using an example from Australia, the paper highlights how qualitative methods can be used to assess the performance of regional development programmes. The paper attempts to provide a possible different way of thinking about how the performance of regional development interventions might be assessed. Comments about their possible application in the European context are provided at the end of the paper.

ISSUES FACING PROGRAMME EVALUATION

Many programmes facilitate the provision of a broad range of social benefits that were not necessarily the core purpose of the programme, and frequently there can be many other unanticipated spin-off benefits as well. Collectively these may contribute significantly to the perceived success of the programme, especially by programme beneficiaries. Should evaluation consider the success of a programme 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 programme 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 (ESTEVEZ and VANCLAY, 2009; JOÃO *et al.*, 2011; VANCLAY, 2003).

Another issue is that an evaluation cannot simply measure whether goals (that is, 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/region anyway? In the field of social impact assessment, the concept of baseline is extended beyond being a single data point fixed in time. Instead, it is argued that the meaningful comparison is not time ($t + 1$) against time (t_0); but rather a comparison at a point in time against what would have happened without the programme (ESTEVEZ *et al.*, 2012). The baseline is thus the line (not point) of expected trending without the programme. In European policy circles, this is called the 'counterfactual' (EVALSED, 2012) – a term borrowed from psychology where it has a slightly different meaning, that is, a mental representation or image of an alternative trajectory, past or future (ROESE, 1997). Thus, programmes can still be regarded as 'successful' if an indicator at a future time is worse than it was at commencement, providing that there is a reasonable analysis to establish that there were other changes taking place such that the programme made the community better off than it would have been without the programme.

In considering a wide range of outcomes, and with the realization that many of the broader social benefits of programmes are subjective, the adage normally attributed to Albert Einstein that 'not everything that counts can be counted' becomes important (VANCLAY, 2012b). Particularly in cases of the enhanced well-being type of programmes 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 routine 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 programme? A short and simple answer might be that they cannot. A more complex answer questions whether simplistic assumptions of direct cause and effect are appropriate. Big programmes with high-level outcomes do not have simple cause-and-effect relationships; they have complex interconnecting multi-causal 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 non-linear and exponential effects) between system elements.

It is important to realize that these debates have existed in the field of evaluation for decades (for a discussion on the purpose and history of evaluation, see GREENE, 2000). While some evaluators have attempted to persist with ever-improving and ever-more-sophisticated empirical quantitative techniques (for example, LEEUW and VAESSEN, 2009), many other evaluation experts fundamentally disagree that such methods can address the complexity of the programmes being considered (for example, 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 favour of qualitative methods (especially narrative approaches) is that they can yield powerful stories that are not only useful for media reports, but also 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 has 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 (that is, convincingly) than other forms of evidence (DENNING, 2007; FISHER, 1989; KURTZ, 2009; LABONTE, 2011; SANDELOWSKI, 1991; SHAW *et al.*, 1998).

A NOTE ON TERMINOLOGY: OUTPUTS AND OUTCOMES

Programme logic refers to the perceived causal understanding of how the different components of a programme (specifically inputs and activities) work together to produce outputs and outcomes. Outcomes are generally conceived as being immediate, intermediate and ultimate (MAYNE, 2004). Programme 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 programme logic (Fig. 1).

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 (for example, dollars spent, hours consumed), activities (for example, numbers of workshops held), and people

involvement (for example, numbers of people attending), arguing that more attention needed to be given to a range of higher-order considerations. Bennett's model, usually represented as a ladder or staircase, went as follows:

- Inputs (resources expended).
- Activities (what was done).
- People involvement (who and how many people involved).
- Reactions (what they thought of it, immediate reactions as might be measured by exit surveys).
- KASA change (that is, changes after a period of time in the knowledge, attitudes, skills and aspirations of participants).
- Practice change (changes in the behaviour of people).
- End results (or ultimate outcomes, that is, what long-term difference did it make).

Bennett's model is not a programme logic model; rather it is a conceptualization of the different stages of a project that should be considered by evaluation. Bennett's argument was that instead of measuring the easy-to-measure things low down in the hierarchy like the inputs, activities and people involvement normally addressed, evaluators should go up the hierarchy and consider all stages. However, 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 programme logic model and a theory of change, some evidence of the likelihood of success can be 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

Fig. 1. Generic depiction of programme logic
Source: MAYNE (1999, p. 9)

evidence may be hard to collect, especially if external conditions are changing, but stories of change from (a selection of) participants can be readily 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.

THE ADDED VALUE OF QUALITATIVE METHODS IN EVALUATION

In general, qualitative research tends (adapted primarily from PADGETT, 2012, and a range of other sources):

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

In evaluation, qualitative research is used specifically to consider the why and how questions that quantitative methods typically cannot answer, for example (adapted from ROYCE *et al.*, 2001; and ANATAS, 2004):

- Why does the programme work (or not work)?
- How does the programme 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 programme developers?
- What were the additional unintended and/or unexpected positive or negative consequences?

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

A SHORT HISTORY OF QUALITATIVE EVALUATION

GUBA and LINCOLN (1989) described four ‘generations’ in the history of evaluation: measurement, description, judgement and constructivist evaluation. Perhaps ‘paradigms’ or ‘modalities’ would have been more appropriate

words to use than ‘generations’, at least in terms of what they mean. A major criticism of this temporal overview of the field of evaluation is that the extent to which the phases (generations) 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/clashes (WADSWORTH, 2005), the dominant culture of evaluation is changing. While the division between summative evaluation and formative evaluation is widely accepted, evaluation is now 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, programme 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 programme 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 amongst all programme and project partners. Evaluation is now seen as a form of action research that informs project and programme design. Ongoing evaluation and adaptive management are essential parts of being innovative and a learning organization. Evaluation is your friend (VANCLAY *et al.*, 2004).

Monitoring and evaluating the performance of rural development programmes 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 programme 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’ (MSC) as the term for the emergent method (DART and DAVIES, 2003; DAVIES and DART, 2005). Since then, the approach has become firmly established in the evaluation and development cooperation professions, and as at July 2013 the term ‘most significant

change technique' had over 52 000 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 MSC. Dart attributes the actual name 'performance story' to John Mayne of the Canadian Auditor General's Office (MAYNE, 2004; DART and MAYNE, 2005). Arguably PSR overcomes some of the criticisms of MSC (WILLETTS and CRAWFORD, 2007) and in particular removes concern about any bias implied by the name and the (mis)perception that the method only looked for positive stories potentially 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 (DART and MAYNE, 2005) and as at July 2013 had over 9000 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 (Jessica Dart, personal communication, 20 April 2011) and because COR emphasizes integrating empirical and qualitative data and does not rely on the story alone. There are a range of other refinements, but these are minor so information about MSC and PSR is essentially still relevant to COR. As at July 2013, 'collaborative outcomes reporting' had over 27 000 hits on Google and had been presented to a number of evaluation conferences.

There are many other story-based or narrative approaches to evaluation that are championed by various people including LABONTE *et al.* (1999), WADSWORTH *et al.* (2007), KURTZ (2009), MAXSON *et al.* (2010) and WITHERS (2010). One of the well-known approaches is that of the Cynefin framework which is implemented in a range of settings with SenseMaker software (KURTZ and SNOWDEN, 2003; SNOWDEN and BOONE, 2007). The US Centers for Disease Control and Prevention (CDCP) advocate a 'success story' approach (CDCP, 2007; LAVINGHOUEZ *et al.*, 2007). Stories and vignettes are also widely used in social research in a range of ways (for example, DARE *et al.*, 2011; HOWDEN and VANCLAY, 2000; HUGHES and HUBY, 2002; SANDELOWSKI, 1991; VANCLAY and ENTICOTT, 2011; ENTICOTT and VANCLAY, 2011).

AN EXAMPLE APPLICATION OF PERFORMANCE STORY REPORTING AT THE PROGRAMME LEVEL

Background to the Australian context

The Australian government funds a range of natural resource management (NRM) programmes and projects

collectively known as 'Caring for Our Country' (<http://www.nrm.gov.au>) (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 programme was about A\$2.25 billion (COMMONWEALTH OF AUSTRALIA, 2008). Investments in the two previous five-year periods totalled A\$1.5 billion and A\$1.3 billion, respectively (AUDITOR GENERAL, 2008). A further A\$2.2 billion has been committed for the five-year period to 2018 (COMMONWEALTH OF AUSTRALIA, 2012).

The programme is currently undergoing considerable change, amidst much criticism (for example, ROBINS and KANOWSKI, 2011; TENNENT and LOCKIE, 2013). It is very broad and supports various efforts to improve NRM across the whole of Australia, depending on the specific needs of each locality/region. It has supported a range of disparate projects often in conjunction with local community (Landcare) groups and individuals (who often have considerable personal investment and commitment to the activities) and may include small grant programmes, co-funding programmes, and support for project staff and project costs. Issues addressed include biodiversity, habitat and remnant vegetation protection and restoration; the establishment of wildlife corridors; water quality and stream health issues; methods to address salinity, acidity and acid sulphate soils and soil erosion; tree-planting activities; weed removal and feral animal issues; and whatever else local regions determine are appropriate and consistent with the government's terms of reference for the programme. PRAGER and VANCLAY (2010) provide a comparison of these land management issues between Germany and Australia. Specific information about the types of activities funded is available from the Australian government's *One Land – Many Stories: Prospectus for Investment* (COMMONWEALTH OF AUSTRALIA, 2012).

While the objectives are now very clear (COMMONWEALTH OF AUSTRALIA, 2008) and there is a clearly elaborated programme logic (ROUGHLEY, 2009), earlier versions of the programme did not have clearly identified intended outcomes. While anecdotal evidence suggested that there was much benefit from the programme (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' (in relation to their obligations in terms of the Commonwealth–State agreements that underpinned the programme), and that:

the quality and measurability of the targets in the regional plans is an issue for attention and ... should be considered nationally – especially as the absence of sufficient scientific data has limited the ability of regional bodies to link the targets in their plans to program outcomes.

This context of official concern about the alleged benefits of the programme, but strong public and political support for the programme led to a real need to prove that the programme was being successful, especially in the knowledge that empirical indicators were unlikely to reveal results in the short-term. It was also an impetus to experiment with new evaluation techniques, since existing methods had failed.

Dart had been experimenting with the MSC and PSR for some time, first in the early to mid-1990s in developing country contexts and later in an agricultural extension context in Australia. Dart's work was known by various people in the Australian government, especially in NRM circles, and it became evident that PSR might be a good way to assist the government in its need to capture the impacts of its NRM investments.

The 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 assessment of the use of story-based approaches (CARR and WOODHAMS, 2008). That feasibility assessment (or 'independent review' as they claimed) considered three questions (CARR and WOODHAMS, 2008, p. 3):

- 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.

Based on four pilot applications of the method, the feasibility assessment concluded that:

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.

(CARR and WOODHAMS, 2008, p. 61)

As a result of the positive feasibility assessment, the PSR approach was rolled out across Australia with the Australian government publishing a 'user guide': *Developing a Performance Story Report* (ROUGHLEY and DART, 2009). In total, around thirteen project-level PSR reports were prepared (the exact number is hard to determine). While not all these reports are publicly available on the internet (for one that is available, see DART and O'CONNOR, 2008), a limited number of hardcopies and PDFs of the reports are in circulation, having been distributed to participants and others who expressed an interest at the time. The PSR method has also been used to assess other programmes and projects, especially in the NRM domain (for example, CLEAR HORIZON and ENVIRONMENT VICTORIA, 2010).

Description of how to undertake a performance story report evaluation

The Australian government's *User Guide* (ROUGHLEY and DART, 2009) and various other instruction manuals (for example, DART *et al.*, 2000; DAVIES and DART, 2005; SILVER *et al.*, 2009) provide instructions on conducting evaluations using PSR. ROUGHLEY and DART (2009, p. 12) suggest that a typical report is between ten and thirty 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 or planning meetings are held to determine what will be evaluated, to develop the programme logic (if not already existing), set evaluation questions, and identify sources of existing evidence and possible people to interview.
- Evidence gathering – an 'evidence trawl' is conducted to identify existing data that will provide the best evidence of expected outcomes. This is followed by the social inquiry process, where interviews are conducted with people who can provide additional information about programme outcomes. Specific questions are asked and recorded to provide stories of significant changes that have occurred as a result of the programme.
- Integrated data analysis – quantitative and qualitative data are analysed to identify evidence corresponding to the outcomes in the programme 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 programme is making towards its stated outcomes. The panel should also identify any further evidence that might be needed to make a conclusive statement about the achievement of programme 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 programme logic to inform a draft set of recommendations.
- Summit meeting – all evaluation participants come together to consider and discuss the findings, nominate the stories that best illustrate the impact of the programme, and make recommendations.
- Integration, report and communications – the evaluator prepares the PSR report, which is a synthesis of all the above steps including the recommendations from summit meeting participants. The findings of the evaluation are communicated to all stakeholders.
- Revising the programme logic – programme managers, staff and other stakeholders meet to consider the report and to revise the programme logic and implementation accordingly.

The *User Guide* (ROUGHLEY and DART, 2009) describes the process of selecting and interviewing respondents. It suggests that programme (project) participants, strategic informants (such as project managers, programme coordinators) and scientific experts who are in a position to comment about likely changes be interviewed. It suggests that depending on the context, between twenty and fifty people need to be interviewed 'to gain an in-depth view of how participants have experienced the program' (p. 23). The *User Guide* also describes the social inquiry process suggesting that questions should focus on outcomes and stories of change. Ultimately the interviewer/evaluation team construct narratives/vignettes/performance stories based on the answers to the questions asked. The interviewer has a responsibility to ensure that the right questions have been asked to allow the participant to provide the information to construct the stories. The stories are in the 'voice' and words of the participants as much as possible, but are edited to ensure that the story is an effective way of explaining what happened. As with all social research, the stories (transcripts) should be given back to the participant for them to approve and edit as they see fit. This process of constructing narratives/vignettes by modifying the actual transcript but remaining true to the intent is not uncommon in social research (for other examples, see DARE *et al.* 2011; and VANCLAY and ENTICOTT, 2011). Three or so stories that best illustrate 'an important change that has occurred as a result of the program' (ROUGHLEY and DART, 2009, p. 46) are selected in a participatory process in the summit meeting (the fifth step). The summit meeting participants also discuss the main issues that emerged and develop recommendations.

The project-level PSR reports can be aggregated into a higher level evaluation or meta-evaluation. One example of this is the assessment of NRM outcomes in the State of South Australia for the period 2001–2008, which drew on several PSR reports that were conducted in that state (DEPARTMENT OF WATER, LAND AND BIODIVERSITY CONSERVATION (DWLBC), 2009).

While the Australian government's *User Guide* makes specific recommendations as to how PSR reports should be presented, arguably they could be presented in a variety of formats, including as conventional consultancy reports, websites, DVDs, posters, brochures, or as mixed media with video and/or audio files. One of the more creative PSR reports contained an appendix comprising a newsletter that had been widely circulated. The newsletter was comic book style, but instead of caricature drawings used real photographs of people in outdoor settings with speech bubbles indicating their name and role which were hot-linked to a website where the reader could listen to an audio file recording of that person's comments (BESSEN CONSULTING SERVICES, 2009).

An assessment of the Australian experiment

As far as can be determined, there has been no publicly available, official, ex-post assessment of the effectiveness of the PSR method or an overall summary of the state of NRM in Australia based on the thirteen or so PSR case studies that were claimed to have been undertaken. Instead, in 2012 the Australian government initiated a completely separate review of the Caring for Our Country programme containing no (or very little) mention of the earlier PSR experiment. The PSR reports that are accessible on the internet in general give a positive impression of both the method and the particular project being evaluated. It is reasonable to suspect that the PSR reports that are not readily accessible were either not favourable towards the specific NRM project being evaluated, and/or demonstrated problems with the application of the method such that the results would be contestable. At some locations, there was cynicism by some members of the expert panels and others with the process. There was also concern by the regions about how time consuming the process was.

The government has since made many changes to the Caring for Our Country programme which have led to much criticism from a range of commentators (for example, ROBINS and KANOWSKI, 2011; TENNENT and LOCKIE, 2013). Whether the non-carry forward of the performance story approach was because of inherent problems with the method or because of the way it was implemented (see below) – or conversely because of the fact that it may have revealed things that the government did not want to hear or to be told – cannot be adequately determined from available sources. It is important to appreciate that the time period in question was politically delicate in Australian politics, particularly with a minority government holding power only with the support of a few independent politicians or minor parties, and a disposing of a sitting prime minister by his (and again later her) own party. What is interesting to note is that while there has been a marked reduction of interest in PSR at the federal level, many local-level agencies have become strongly committed to PSR and continue to produce PSR reports. Therefore, reasonable academic reflection on what has transpired in Australia would suggest that the fact that the government has not continued with PSR is not indicative of a problem with the method.

For the purposes of this paper, it does not actually matter what ultimately happened in Australia. The Australian case is being presented as proof that qualitative evaluation can be applied in practice. While there may be areas for possible improvement with the method (see discussion below), these do not fundamentally detract from its potential value in Australia, in Europe, or elsewhere. Qualitative evaluation methods are being outlined in this paper as being of potential use (something to consider), and this paper is not necessarily

specifically recommending the PSR approach in exactly the way implemented in the Australian context. It does however recommend the approach in general.

CRITICAL REFLECTIONS ON THE PERFORMANCE STORY REPORTING METHOD

The commentary provided here draws on a cursory review of the PSR reports available to the author, discussions with various participants and stakeholders in the PSR approach and NRM in Australia, and the author's personal experience in participating in a summit meeting. The author, an experienced social research methodologist and rural sociologist/geographer, has been an observer of NRM in Australia for over three decades (for example, VANCLAY, 2004). He relocated from Australia to the Netherlands in 2010. While there is no claim that this is a thorough review, the observations below are likely to be indicative of the issues that would need to be considered in any further utilization of the method.

In comparison with the MSC, in PSR ironically the role of stories has been downplayed, and the importance of programme 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 the case in COR, the next incarnation of the method. It is likely that there was scepticism about the 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 (that is, 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. However, a problem arises when the expert panel does not fundamentally believe in the process, or when their views are markedly different to those of the project participants. Arguably, the stakeholder panel alone should be a sufficient validation process.

Curiously, PSR was used in Australia because there was no pre-existing programme logic and little capacity to utilize data other than the performance stories. The unique value of the method which led to its use was its ability to be a retrospective assessment and to provide some evidence of change in the absence of baseline data. The stories of participants gave a sense of the multifaceted nature of the outcomes, were able to adjust for the confounding effects of external events (such as the drought most of Australia experienced for much of the previous decade), and detailed the unanticipated benefits that were experienced.

Anecdotal evidence suggests that involvement in the PSR process was very rewarding for participants. They learned things about what they were doing that they had not thought of before. This enabled them to consider this aspect of their activities and to enhance upon it. The programme evaluation was also done after an extended dry period, when many farmers and rural residents were hurting and many NRM schemes (tree planting etc) were not particularly effective. Instead of the evaluation process proving that the schemes were yet another failure leading to further disappointment (and mental health deterioration) to participants (HUNT *et al.*, 2011), the evaluations revealed to the participants that there were benefits from these programmes anyway. Thus, it was a valuable self-validation process. Project managers and NRM agency staff also mentioned that they learned about the factors and inputs that contributed to change and outcomes in ways that survey-type evaluations would not have achieved. And because of the participatory process, these findings were self-realizations (personal learnings) rather than the conclusions of some external expert which would not necessarily be accepted by the audience. As mentioned above, many managers at the regional level have become strong supporters of the PSR approach and continue to use it in their agencies, despite lack of interest at the Australian federal level.

There is no doubt that the majority of participants in the process thought it valuable and worthwhile. In general, they thought that the story approach was refreshing and that it validated their personal views. Bureaucrats, especially those not familiar with qualitative research methodologies, were unimpressed. They did not accept that the stories were evidence of outcome, and they were worried about a potential bias in the selection of stories favouring atypical positive stories. The implication of this observation is that there is no point in using the PSR method, unless process work is done with key stakeholders to ensure that they understand how the method works and so that they accept the legitimacy of the approach.

While the use of the method was arguably successful in the Australian context (despite its later abandonment), there were some issues that could be addressed should the methodology be used again in the future in Australia or elsewhere. The PSR method has some inherent limitations. For example, 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 (for example, 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 (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 PSR 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 professionalism 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 European Union Cohesion Funds, especially as an augmentation to other forms of evaluation.

SPECULATION ON THE FEASIBILITY OF STORY-BASED EVALUATION IN THE CONTEXT OF EUROPEAN UNION COHESION POLICY

The application of a story-based approach to evaluation in the Australian NRM context establishes the potential of the approach. There is nothing particularly unusual about the Australian situation that would mean that it would apply there but not elsewhere such as in Europe or North America. Indeed, it has been used in many other contexts, including in developing countries (WILLETTTS and CRAWFORD, 2007), in Canada where it was used to evaluate the effectiveness of the various cultural programmes undertaken by the City of Surrey as part of its Cultural Capital of Canada Award

(WITHERS, 2010), and in the USA where a similar approach is promoted by the US Department of Health through the CDCP (2007). Thinking about the European situation and the types of interventions typically funded suggests the following observations may be appropriate.

- Local development interventions in an urban or rural setting – this seems to be an ideal situation for the use of story-based evaluation. Because of the disparateness of the range of activities, selecting common indicators could be difficult. Story-based evaluation would enable the collation of evidence of change even where the on-ground activities varied considerably.
- Innovation support and/or enterprise support – companies and other organizations (for example, universities) could create stories about what they used the support for and what difference the support made. In these cases, empirical indicators may be available – at least in terms of the improvement in the financial performance of profit-oriented companies; however, the stories are likely to highlight additional added-value dimensions that are not revealed through the numbers alone.
- Support for dynamic systems involving numerous different stakeholders such as clusters – measurement in such contexts is always difficult because of differing units of analysis. The dynamic nature implies that the system is in a state of flux such that movement on any one indicator does not necessarily explain what is happening. Stories have a greater potential to explain how the funding support made a difference to different stakeholders in the system.
- Incentive grants and loans intended to stimulate behaviour change – potentially these initiatives are difficult to evaluate with existing indicators because they are often quite specific. Nevertheless, often ad hoc indicators can be developed that measure the changed behaviour, especially where the behaviour relates to consumption. However, where the behaviour change is not immediate, it may be harder to identify appropriate indicators. Here story-based evaluation can assist. Qualitative approaches might also provide a greater understanding about why the programme led to changed behaviour. Perhaps the grant or loan was not the main cause and the measured relationship was spurious. Qualitative evaluation would also be able to provide a sense of whether the size of the grant or loan was appropriate to have the necessary stimulus effect, or whether it needed to be increased to expand uptake, or whether it could be decreased without sacrificing its effectiveness.
- Training and capacity building – measures of investment for training and capacity building can readily assess the numbers of people attending activities, and where there are assessment (examination) processes, the number who passed. Satisfaction surveys

can also be undertaken. However all these measures are low down on Bennett's hierarchy, that is, they are largely measures of inputs, activities, outputs, or satisfaction and are not adequate measures of ultimate, intermediate or even immediate outcomes. The outcomes of training should focus on what difference the training makes to the lives of those who did the training. Story-based evaluation is likely to provide much more information about the outcomes of training and capacity building programmes than any empirical measure can.

- Investment in infrastructure (roads, rail, environmental infrastructure) – it is worth highlighting that story-telling approaches might reveal many more benefits (and potentially problems) about improvements in infrastructure. Improved public transport, for example, not only reduces journey-to-work time, but by making it accessible to mobility-restricted people potentially makes a world of difference to them enabling them to get jobs, have a wider range of entertainment, better access to shopping, and greater autonomy over their lives. The crude empirical indicators give no sense of the richness and value that increasing transport options can have.

Thus, in all types of European Union funding, whether or not empirical indicators are available, story-based approaches will always provide additional information. In a summative context, story-based approaches will provide additional evidence especially of the extra 'social return on investment'. In a formative context, story-based and qualitative evaluation will provide more information about how the programme can be improved.

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; stories are more meaningful; stories are more real; stories convey information more effectively and are more likely to be remembered than facts.

An effective story, however, has to be a proper story. It cannot 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. It needs to be multidimensional, but the different components need to be interconnected 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 behaviour 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, analyse and make sense of things in a formal way is relatively recent, but has been part of strategic planning in business for some decades. Using stories as an evaluation methodology is more recent, but has much appeal and, as demonstrated by the Australian experience, has the potential to 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 additional meaning and interpretation. Big programmes are subject to long lag times, and can be subject to the influence of multiple external influences. It is naïve to think that having 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.

Acknowledgements – The document on which this paper is based was prepared at the request of the European Commission's Directorate-General for Regional Policy. Significant comments on it and/or drafts of this paper and their associated presentations were received from (in alphabetical order): Fabrizio Barca, Veronica Gaffey, Phillip McCann and Marielle Riche. Thanks to Jessica Dart for agreeing to be interviewed. Thanks also to many Australian natural resource management (NRM) stakeholders with whom these issues were discussed. The original document is available from the DG-Regio website at: http://ec.europa.eu/regional_policy/sources/docgener/evaluation/doc/performance/Vanclay.pdf

REFERENCES

- ANATAS 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 Number 21 2007–08). Australian National Audit Office, Canberra, ACT (available at: http://www.anao.gov.au/uploads/documents/2007-08_Audit_Report_21.pdf) (accessed on 11 April 2012).
- BENNETT C. (1975) Up the hierarchy, *Journal of Extension* **13**(2), 7–12.
- BESSEN CONSULTING SERVICES (2009) *Performance Story Report: Evaluation of Investment in the Dugong and Marine Turtle Project*. North Australian Indigenous Land and Sea Management Alliance, Darwin, NT (available at: <http://www.savanna.org.au/nailsma/publications/downloads/Dugong-and-Marine-Turtle-Project-PSR.pdf>) (accessed on 28 April 2013).
- CARR A. and WOODHAMS F. (2008) *Monitoring, Evaluating, Reporting and Improving (MERI) Natural Resource Management: Making MERI in Australia*. Bureau of Rural Sciences, Canberra, ACT.
- 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), 8086–8091.
- CENTERS FOR DISEASE CONTROL AND PREVENTION (CDCP) (2007) *Impact and Value: Telling Your Program's Story*. CDCP, National Centers for Chronic Disease Prevention and Health Promotion, Division of Oral Health, Atlanta, GA (available at: http://www.cdc.gov/oralhealth/publications/library/pdf/success_story_workbook.pdf) (accessed on 28 April 2013).
- CLEAR HORIZON AND ENVIRONMENT VICTORIA (2010) *Regional Sustainable Living Program: Performance Story Report*. Environment Victoria, Melbourne, VIC (available at: <http://environmentvictoria.org.au/sites/default/files/RSL%20Whole%20of%20Program%20Report%20-%20public%20version.pdf>) (accessed on 28 April 2013).
- COMMONWEALTH OF AUSTRALIA (2008) *Caring for Our Country Outcomes 2008–2013*. Commonwealth of Australia, Canberra, ACT (available at: <http://nrmonline.nrm.gov.au/catalog/mql:1887>) (accessed on 28 April 2013).
- COMMONWEALTH OF AUSTRALIA (2012) *One Land – Many Stories: Prospectus for Investment*. Australian Government Department of Sustainability, Environment, Water, Population and Communities, Canberra, ACT (available at: <http://www.environment.gov.au/biodiversity/publications/prospectus/index.html>) (accessed on 25 July 2013).
- CURTIS A. and LOCKWOOD M. (2000) Landcare and catchment management in Australia: Lessons for state-sponsored community participation, *Society and 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), 1149–1168.
- DART J. and DAVIES R. (2003) A dialogical, story-based evaluation tool: the Most Significant Change Technique, *American Journal of Evaluation* **24**(2), 137–155.
- DART J., DRYSDALE G., COLE D. and SADDINGTON M. (2000) *The Most Significant Change Approach for Monitoring an Australian Extension Project*. PLA Notes Number 38, pp. 47–53. International Institute for Environment and Development, London.
- DART J. and MAYNE J. (2005) Performance story, in MATHISON S. (Ed.) *Encyclopedia of Evaluation*, pp. 307–308. Sage, Thousand Oaks, CA.
- 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, ACT (available at: <http://nrmonline.nrm.gov.au/catalog/mql:2167>) (accessed on 11 April 2012).
- 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] (available at: <http://www.mande.co.uk/docs/MSCGuide.pdf>) (accessed on 11 April 2012).
- DENNING S. (2007) *The Secret Language of Leadership: How Leaders Inspire Action through Narrative*. Jossey-Bass, San Francisco, CA.
- DEPARTMENT OF WATER, LAND AND BIODIVERSITY CONSERVATION (DWLBC) (2009) *Our Changing Environment: Outcomes and Lessons from Investment in South Australia's Natural Resources 2001–2008*, Government of South Australia, Department of

- Water, Land and Biodiversity Conservation, Adelaide, SA (available at: <http://nrmonline.nrm.gov.au/catalog/mql:2677>) (accessed on 10 October 2013).
- ENTICOTT G. and VANCLAY F. (2011) The moral accountability of agricultural scripts: an analysis of farmers' talk about animal health risks, *Health Risk and Society* **13**(4), 293–309.
- ESTEVEZ 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.
- ESTEVEZ 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), 137–145.
- EVALSED (2012) *EVALSED: The Resource for the Evaluation of Socioeconomic Development* (available at: http://ec.europa.eu/regional_policy/sources/docgener/evaluation/guide/guide2012_evalsed.pdf) (accessed on 14 September 2013).
- FELLER I. (2007) Mapping the frontiers of evaluation of public sector R&D programs, *Science and Public Policy* **34**(10), 681–690.
- FETTERMAN D. M. (2000) *Foundations of Empowerment Evaluation*. Sage, Thousand Oaks, CA.
- FETTERMAN D. M., KAFTARIAN S. J. and WANDERSMAN A. (Eds) (1995) *Empowerment Evaluation*. Sage, Thousand Oaks, CA.
- FISHER W. R. (1989) *Human Communication as Narration: Toward a Philosophy of Reason, Value, and Action*. University of South Carolina Press, Columbia, SC.
- GREENE J. C. (2000) Qualitative program evaluation: practice and promise, in DENZIN N. K. and LINCOLN Y. S. (Eds) *Handbook of Qualitative Research*, pp. 530–544. Sage, Thousand Oaks, CA.
- GUBA E. and LINCOLN Y. (1989) *Fourth Generation Evaluation*. Sage, Newbury Park, CA.
- 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 and Research Centre, Wageningen (available at: http://capacity.org/capacity/export/sites/capacity/documents/topic-readings/110412-evaluation-revisited-may-2010_small-version.pdf) (accessed on 11 April 2012).
- HAKOWICZ S. (2009) The evolution of Australia's natural resource management programs: towards improved targeting and evaluation of investments, *Land Use Policy* **26**(2), 471–478.
- 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), 382–386.
- HUNT W., VANCLAY F., BIRCH C., COUTTS J., FLITTNER J. and WILLIAMS B. (2011) Agricultural extension: Building capacity and resilience in rural industries and communities, *Rural Society* **20**(2), 112–127.
- JACKSON E. T. and KASSAM Y. (Eds) (1998) *Knowledge Shared: Participatory Evaluation in Development Cooperation*. Kumarian, West Hartford, CT.
- JOÃO E., VANCLAY F. and DEN BROEDER L. (2011) Emphasising enhancement in all forms of impact assessment, *Impact Assessment and Project Appraisal* **29**(3), 170–180.
- KURTZ C. (2009) *Working with Stories in your Community or Organization* (available at: <http://www.workingwithstories.org>) (accessed on 11 April 2012).
- KURTZ C. and SNOWDEN D. (2003) The new dynamics of strategy: sense-making in a complex and complicated world, *IBM Systems Journal* **42**(3), 462–483.
- LABONTE R. (2011) Reflections on stories and a story/dialogue method in health research, *International Journal of Social Research Methodology* **14**(2), 153–163.
- LABONTE R., FEATHER J. and HILLS M. (1999) A story/dialogue method for health promotion knowledge development and evaluation, *Health Education Research* **14**(1), 39–50.
- LAUGHLIN R. and BROADBENT J. (1996) Redesigning fourth generation evaluation, *Evaluation* **2**(4), 431–451.
- LAVINGHOUEZ R., PRICE A. and SMITH K. (2007) The program success story: a valuable tool for program evaluation, *Health Promotion Practice* **8**(4), 323–331.
- LEEUEW F. and VAESSEN J. (2009) *Impact Evaluations and Development: NONIE Guidance on Impact Evaluation. The Network of Networks on Impact Evaluation*. c/- Independent Evaluation Group, The World Bank, Washington, DC (available at: http://siteresources.worldbank.org/EXT/OED/Resources/nonie_guidance.pdf) (accessed on 11 April 2012).
- MAXSON M., GUIJT I., et al. (2010) *The 'Real Book' for Story Evaluation Methods*. GlobalGiving Foundation (available at: <http://www.globalgiving.org/jcr-content/gg/landing-pages/story-tools/files/-story-real-book--2010.pdf>) (accessed on 11 April 2012).
- MAYNE J. (1999) *Addressing Attribution through Contribution Analysis: Using Performance Measures Sensibly (Discussion Paper)*. Office of Auditor-General of Canada (available at: <http://dsp-psd.pwgsc.gc.ca/Collection/FA3-31-1999E.pdf>) (accessed on 11 April 2012).
- MAYNE J. (2004) Reporting on outcomes: setting performance expectations and telling performance stories, *Canadian Journal of Program Evaluation* **19**(1), 31–60.
- PADGETT D. K. (2012) *Qualitative and Mixed Methods in Public Health*. Sage, Thousand Oaks, CA.
- PATTON M. Q. (1981) *Creative Evaluation*. Sage, Beverly Hills, CA.
- PATTON M. Q. (1990) *Qualitative Evaluation and Research Methods*. Sage, Newbury Park, CA.
- PATTON M. Q. (1997) *Utilization-Focused Evaluation*. Sage, Thousand Oaks, CA.
- PRAGER K. and VANCLAY F. (2010) Landcare in Australia and Germany: comparing structures and policies for community engagement in natural resource management, *Ecological Management and Restoration* **11**(3), 187–193.

- ROBINS L. and KANOWSKI P. (2011) 'Crying for our Country': eight ways in which 'Caring for our Country' has undermined Australia's regional model for natural resource management, *Australasian Journal of Environmental Management* **18**(2), 88–108.
- ROESE N. (1997) Counterfactual thinking, *Psychological Bulletin* **121**(1), 133–148.
- ROUGHLEY A. (2009) *Developing and Using Program Logic in Natural Resource Management: User Guide*. Australian Government, Canberra, ACT (available at: <http://nrmonline.nrm.gov.au/catalog/mql:2164>) (accessed on 11 April 2012).
- ROUGHLEY A. and DART J. (2009) *Developing a Performance Story Report. User Guide*. Australian Government, Canberra, ACT (available at: <http://nrmonline.nrm.gov.au/catalog/mql:2162>) (accessed on 11 April 2012).
- ROYCE D., THYER B. T., PADGETT D. K. and LOGAN T. K. (2001) *Program Evaluation: An Introduction*, 3rd Edn. Brooks/Cole, Pacific Grove, CA.
- SANDELOWSKI M. (1991) Telling stories: narrative approaches in qualitative research, *Image: Journal of Nursing Scholarship* **23**(3), 161–166.
- 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, VIC (available at: <http://lwa.gov.au/files/products/reporting/pn30209/pn30209.pdf>) (accessed on 11 April 2012).
- SNOWDEN D. and BOONE M. (2007) A leader's framework for decision making, *Harvard Business Review* **85**(11), 69–76.
- TENNENT R. and LOCKIE S. (2013) Vale Landcare: the rise and decline of community-based natural resource management in rural Australia, *Journal of Environmental Planning and Management* **56**(4), 572–587.
- VANCLAY F. (2003) International principles for social impact assessment, *Impact Assessment and Project Appraisal* **21**(1), 5–11.
- VANCLAY F. (2004) Social principles for agricultural extension to assist in the promotion of natural resource management, *Australian Journal of Experimental Agriculture* **44**(3), 213–222.
- VANCLAY F. (2012a) *Guidance for the Design of Qualitative Case Study Evaluation: A Short Report to DG Regio* (available at: http://ec.europa.eu/regional_policy/sources/docgener/evaluation/doc/performance/Vanclay.pdf) (accessed on 17 May 2012).
- VANCLAY F. (2012b) The potential application of Social Impact Assessment in coastal zone management, *Ocean and Coastal Management* **68**, 149–156.
- VANCLAY F. and ENTICOTT G. (2011) The role and functioning of cultural scripts in farming and agriculture, *Sociologia Ruralis* **51**(3), 256–271.
- 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 and Management* **6**(4), 539–564.
- WADSWORTH Y. (2005) 'Gouldner's child?': some reflections on sociology and participatory action research, *Journal of Sociology* **41**(3), 267–284.
- 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, Melbourne, VIC.
- 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 available at: <http://www.rh.org.au/articles/subviewnew.asp?ArticleID=1555>) (accessed on 11 April 2012).
- WHITMORE E. (Ed.) (1998) *Understanding and Practicing Participatory Evaluation: New Directions for Evaluation*. Jossey-Bass, San Francisco, CA.
- WHYTE W. F. (Ed.) (1990) *Participatory Action Research*. Sage, Newbury Park, CA.
- WILDER L. and WALPOLE M. (2008) Measuring social impacts in conservation: experience of using the Most Significant Change method, *Oryx* **42**(4), 529–538.
- WILLETTTS J. and CRAWFORD P. (2007) The most significant lessons about the Most Significant Change Technique, *Development in Practice* **17**(3), 367–379.
- WITHERS D. (2010) *Narrative Evaluation: City of Surrey's Cultural Capital of Canada Award Program* (available at: <http://denisewithers.files.wordpress.com/2011/02/surreycccereport1.pdf>) (accessed on 28 April 2013).