

EJA

EVALUATION
Journal of Australasia

Vol 15 | No 4 | 2015



Edited by
Lyn Alderman
Janet Clinton

15 | 4



aes 16

International **Evaluation** Conference

17–21 September 2016 Perth Australia

'Landscapes'



Call for papers

Opens 14 December 2015

Visit the conference website for
more information.



conference2016.aes.asn.au

Contents

2

Editorial

4

REFEREED ARTICLE

Context-sensitive evaluation: determining the context surrounding the implementation of a government policy

Lyn Alderman

16

REFEREED ARTICLE

System evaluation theory (SET): a practical framework for evaluators to meet the challenges of system evaluation

Ralph Renger

29

REFEREED ARTICLE

Building evaluation capacity in micro community organisations —more burden than benefit?

Bridget Doherty, Richard Eccleston, Emily Hansen, Kristin Natalier, Brendan Churchill

38

PRACTICE ARTICLE

Reflexive monitoring in New Zealand: evaluation lessons in supporting transformative change

Kelly Rijswijk, Denise Bewsell, Bruce Small, Paula Blackett

44

BOOK REVIEW

Title: *Interviewing for Qualitative Inquiry: A Relational Approach*
reviewed by *Ruth Davey*

46

Notes to contributors

Edited by
Lyn Alderman
Janet Clinton

Book reviews edited by
Darrel N Caulley

Assistant Editors

Dr Carol Quadrelli
Dr Liz Gould
Dr Bronwyn Rossingh

The *Evaluation Journal of Australasia* is published quarterly and is the official journal of the

Australasian Evaluation Society Ltd
PO Box 476, Carlton South VIC 3053
Australia

Tel: +61 3 9035 3469

Email: aes@aes.asn.au

Website: www.aes.asn.au

Intending contributors should consult the 'Notes to Contributors' (p. 46)

Join the AES at: www.aes.asn.au

Institutional subscriptions available at
<http://informit.com.au>

Copyright © 2015 Australasian Evaluation Society Ltd and individual contributors

The responsibility for all statements made in the *Evaluation Journal of Australasia* lies with the authors. The views of the authors included in the journal do not necessarily reflect the views and position of the editors or the AES. The AES does not endorse advertised products and services.

Copy editing:

Rachel Aston

Cover, page design and layout:

Ingrid Ciotti (Studio 673)

ISSN 1035-719X

Editorial

Lyn Alderman | Janet Clinton | *Editors*

Welcome everyone to our fourth and final issue of the Evaluation Journal of Australasia (EJA) for 2015—the ‘Year of Evaluation’.

Before discussing the content for this issue, we would like to bring your attention to four milestones that have made a significant contribution to the profile of the journal. These milestones consequently place the EJA in a good position to apply for inclusion in two academic online databases. The editorial team has worked diligently to prepare the journal to meet the application criteria for inclusion within Elsevier’s Scopus (the largest abstract and citation database of peer-reviewed literature including scientific journals, books and conference proceedings). A subsequent application will be submitted to Thomson Reuters’ Web of Science Core Collection (a tool that provides researchers, administrators, faculty and students with quick, powerful access to the world’s leading citation databases).

Milestone 1:

In September 2014, the editorial team made the commitment to move the EJA from a print to an online format. The purpose of this was to revitalise the format of the journal and enable an increase in issues—from biannual to quarterly. The new format and frequency are both firsts in the history of the journal.

Milestone 2:

The publication of this issue will mark the sixth consecutive issue published, on schedule, over an 18-month period, thus making EJA worthy of inclusion in international databases. To ensure that the issues are published on schedule and to maintain the standard of the journal, the editorial team adopted a minimum issue format—including an editorial, three peer reviewed articles, a newly introduced practice article and a book review. In support of this format and to subsequently increase content, author guidelines between the AES conference and the EJA have been aligned more closely.

Milestone 3:

Following advice from the Australasian Evaluation Society (AES) Publications and Journal Committee, an invitation was sent out to all members inviting interested parties to submit an expression of interest for one of three Assistant Editor positions. Of the ten applications received, it is our pleasure to announce that Dr Carol Quadrelli, Dr Liz Gould and Dr Bronwyn Rossingh have accepted a three-year voluntary position as Assistant Editors of the EJA. These appointments place the editorial team in a great position for succession planning.

Milestone 4:

Understanding and applying ethical practice to the publication of authentic material is an ongoing imperative. As evaluators and members of the AES, ethics is a fundamental element of evaluation theory, practice and use; all AES members are obliged to uphold the *AES Code of Ethics* and apply the *AES Guidelines for the Ethical Conduct in Evaluations*. When reviewing the publication framework that outlines the guidelines for authors and processes for the review of articles, the editorial team found that further clarity around the roles and responsibilities for editors, assistant editors, reviewers and authors was needed. Following an environmental scan on ethics in publishing, the editorial team located the International Committee on Publishing Ethics online at <http://publicationethics.org/> and determined that the guidelines of the Committee to be an appropriate mechanism to guide the EJA into the future. Therefore, the editorial team has adopted the International Committee on Publishing Ethics’ *Code of Conduct for Journal Editors* and *Code of Conduct and Best Practice Guidelines for Journal Editors*; these guidelines provide detailed general duties and responsibilities pertaining to an editor’s relationships with readers, authors, reviewers, and publishers.



In providing detail about the governance of EJA, the editorial team would like to reassure our AES members and broader readership that the governance of this journal is well in hand and that we take a very serious, deliberate and mindful approach to our stewardship of the EJA.

In this issue of the EJA, the articles offer researchers and evaluators an opportunity to explore interrelated topics including: context-sensitive evaluation; the application of system evaluation theory; the importance of evaluation capacity building to support community service implementation; and, the value of system-based reflexive monitoring.

Alderman's article, 'Context-sensitive evaluation: determining the context surrounding the implementation of a government policy' demonstrates to evaluators the significance of context. In an initial study, the application of Rog's model of contextual parameters to the implementation of a government policy, provided the evaluator with an opportunity to unpack six parameters that had direct bearing on how the broader program evaluation would be examined. Specifically, this process redefined the scope of the broader program; identified situations where the political context could directly effect the policy implementation (regardless of whether the implementation was successful or not); confirmed a gap within the literature that substantiated the need for the broader program evaluation; and, identified the importance of 'grey' literature as an essential component.

Renger's article, 'System evaluation theory (SET): a practical framework for evaluators to meet the challenges of system evaluation' will challenge all evaluators to step back from the practice of program evaluations to a broader, more complex evaluation of systems. After setting the background context, this article addresses a gap in the literature and provides a set of three guiding principles to assist evaluators to approach SET in a systematic manner. Of particular interest, is the manner in which the article addresses the issue of how to demystify the complexities of multiple program evaluations that are inter-related, culturally-specific and dependent on the information technology available at the time. Evaluators who are interested in delving into systems thinking, practice and use will find this a great read.

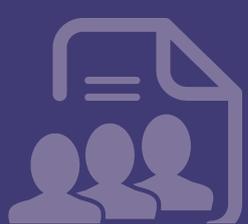
Doherty, Eccleston, Hansen, Natalier and Churchill's article, 'Building evaluation capacity in micro-community organisations—more burden than benefit?' documents the difficulties that not-for-profit and community organisations face when introducing evaluation at the service delivery level. In providing the background literature to the origins of the evidence-based policy movement and explaining the reasons why capacity building is useful, this article goes on to outline the issues raised in community housing projects in Tasmania. The article presents challenges in the use of case study methods such as geographical dispersion; part-time staff with varied levels of experience; limited service funding; and, consequently, limited time for implementation. These challenges led the authors to question whether evaluation capacity building at the community organisation service delivery level was 'more burden than benefit'.

Rijswijk, Bewsell, Small and Blackett's practice article, 'Reflexive monitoring in New Zealand: evaluation lessons in supporting transformative change' offers a practical view of an action research approach to evaluating a system. The system in this context is the New Zealand Government's approach to change, through the implementation of an innovative agricultural system. The evaluation adopted a reflexive monitoring approach; this required the employment of a reflexive monitor for each project. There is merit for evaluators to consider reflexive monitoring whereby expert guidance is provided along the way—being mindful that the expert guidance may require capacity building activities for a successful outcome.

Davey's review of the book, *Interviewing for Qualitative Inquiry: A Relational Approach* offers a practical guide to assist researchers and evaluators to develop a structure to their interviews 'to elicit a narrative, which responds directly to the research question'. This book reinforces the importance of ethical behaviour and practice that will support good research and evaluation outcomes.

In closing, the editorial team would like to thank the authors, reviewers and readers for their ongoing support of the EJA and to wish seasons greetings to all. Thank you to everyone for supporting the 2015 Year of Evaluation and we look forward to what 2016 may bring.



**LYN ALDERMAN**

Context-sensitive evaluation: determining the context surrounding the implementation of a government policy

This article explains the essence of the context-sensitive parameters and dimensions in play at the time of an intervention, through the application of Rog's (2012) model of contextual parameters. Rog's model offers evaluators a structured approach to examine an intervention. The initial study provided a systematic way to clarify the scope, variables, timing, and appropriate evaluation methodology to evaluate the implementation of a government policy. Given that the government implementation of an educational intervention under study did not follow the experimental research approach, nor the double cycle of action research approach, the application of Rog's model provided an in-depth understanding of the context-sensitive environment; it is from this clear purpose that the broader evaluation was conducted. Overall, when governments or institutions implement policy to invoke educational change (and this intervention is not guided by an appropriate evaluation approach), then program evaluation is achievable post-implementation. In this situation, Rog's (2012) model of contextual parameters is a useful way to achieve clarity of purpose to guide the program evaluation.

Introduction

Historically, social programs have been examined through traditional experimental research (Gall, Gall & Borg 2005). A study conducted in 1932 with an experimental and a control group looked at differences between those who used manual typewriters for writing, and those who used traditional pencils and paper. Two thousand participants from 200 schools were involved in the study, that found that overall the use of technology did not directly impact the development of composition skills among participants (Bugelski 1971; Heafner 1932; Wood & Freeman 1932). Although this study supported the researchers' hypothesis, these types of educational experiments came under harsh criticism in the 1960s (Weiss 1983). It was argued that educational experiments disadvantaged the participants within the control group, took a long time to enact and provided little evidence of long term impact. As a result, innovative educational

program evaluations were expanded to be inclusive of all participants in their real world contexts and thus, came to be known as interventions (Gall et al. 2005).

With the movement away from educational experiments and towards innovative educational program evaluations, a different research approach was required. One example of a different approach can be found in action research, which emerged in the 1940s. Kemmis and McTaggart (1988), well-known authors of the *Action Research Planner*, strongly supported action research as a way to examine the benefits of changes in education that was inclusive of all participants. This process generally involved six steps: (i) planning a change; (ii) acting and observing the process and consequences of the change; (iii) reflecting on these processes and consequences; (iv) re-planning; (v) acting and observing again; and (vi) reflecting again (Kemmis, McTaggart & Retallick 2004). Nevertheless, this double cycle of



Lyn Alderman is Associate Director of Academic Quality and Standards in the Learning and Teaching Unit, Queensland University of Technology.
Email: lyn.alderman@qut.edu.au

action, while effective in educational program evaluations, was not always practiced when it came to implementing government policy designed to bring about change in educational practice.

Within the management literature, the term ‘implementation analysis’ is used when a researcher is interested in determining outcomes of the implementation of a policy (Ryan 1999). From an evaluation perspective, this type of investigation is called program evaluation, particularly in education, where policy deployment effectively works as an intervention (Owen & Rogers 1999). For the purposes of this evaluation, the term ‘program evaluation’ will be used, as it strongly aligns with evaluation research. In particular, program evaluation is intended to assist decision-makers (in this case, the Australian Government) to make a record of times and events, and provide a useful guide to the future (Kogan 2005). Decision-makers are being asked to plan more carefully and reflect more critically in order to justify the decisions made (Owen & Rogers 1999). However, program evaluation is often linked to policy borrowing in education, where a policy is borrowed from one setting and implemented in another.

For instance, there is a body of literature that raises concerns that when planning a change in policy, governments often borrow policies from elsewhere, regardless of whether the policy was successful in its original context and with little consideration of what adaptations may be required for application in its new

context (Lingard 2010; Lingard & Garrick 1997). If a policy is borrowed and implemented without sufficient modification for the local context, it may subsequently be found to be unsuccessful, or even deemed to be a ‘managerial fad’ (Birnbaum 2000; Ponzi & Koenig 2002); these issues are covered in more depth in Alderman (2015).

Thus, the problem occurs when a government or institution engages in the implementation of a new policy or educational program evaluation, without the benefit of a planned evaluation approach—such as ‘double cycle of action research’ or a ‘planned evaluation’. Such action research or evaluation would take into account the broader context in which the policy implementation occurs. It therefore follows that to evaluate the success of an intervention, the context surrounding the initial implementation is critical (Rog 2012). It is this very context that may offer vital information that will lead the evaluator to identify enablers or blockers that may impact directly on the outcomes of the intervention.

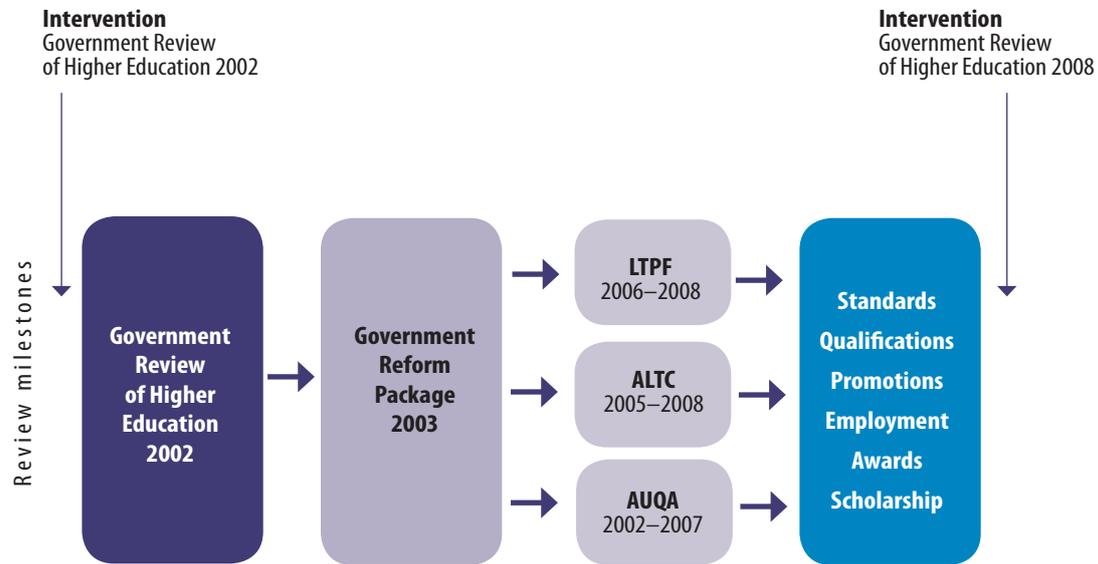
This article aims to demonstrate the benefit of careful planning of a program evaluation that elicits the context-sensitive parameters and dimensions at play at the time of the intervention. Further, it aims to unpack the multiple layers of context that were in play at the time of the 2003 Australian Government Reform Package (Nelson 2003). As Rog (2012) argued, understanding context is an important dimension of any evaluation and will result in a descriptive account of the state of play at the time of the intervention. However, when conducting program evaluation, it is imperative to provide the full story to understand the parameters or variables that may influence the outcomes (Maggetti, Gilardi & Radaelli 2012); this will be achieved by describing how the application of Rog’s model of contextual parameters (Figure 2, p. 7) to an intervention can provide clarity of purpose to the broader program evaluation.

More specifically, this article explains an initial study that sits within a broader program evaluation (Alderman 2014); it evaluates the influence of three national learning and teaching initiatives that emerged from the 2003 Australian Government Reform Package in Higher Education (Nelson 2003). The broader program evaluation represents an independent examination of the 2002 Government Review and consequent 2003 Government Reform Package, by adopting a methodology to conduct an evaluation of a large-scale illuminative program evaluation, unique in its focus on learning and teaching. Figure 1 (p. 6) outlines the specific milestones associated with the review under examination, within the broader program evaluation.

As identified in Figure 1, there are three main initiatives of interest within the 2003 Reform Package: (i) the Learning and Teaching Performance Fund (LTPF);



FIGURE 1: MODEL OF CONTEXTUAL PARAMETERS (ROG 2012)



(ii) the Australian Learning and Teaching Council (ALTC); and, (iii) the Australian Universities Quality Agency (AUQA). These initiatives were implemented within an environment of constant change in political leadership, where policies are borrowed and often implemented within a contracted timeframe. Thus, there was a particular and unique set of contextual parameters that were in play at the time of implementation, these contextual parameters form the basis of the method discussed below.

Method

This initial study adopted Rog’s model of contextual parameters (Figure 2) to provide the environmental situation in which a government implemented an educational policy. It is through this model that the 2003 Australian Government Reform Package (Nelson 2003) may be regarded as ‘an intervention’; the model offered an opportunity to understand the specific triggers leading to the intervention and set the scene for investigation in the broader study. As shown in Figure 2, such an intervention sits within multiple contexts, including: (i) the ‘broader environment context’ (within which social or organisational contexts does the issue reside?); (ii) the ‘problem context’ (what is the intervention ostensibly trying to address?); (iii) decision-making (what is the current political context of decision-making?); (iv) the ‘intervention context’ (what is the program supposed to

be?); (v) the ‘evaluation context’ (how was the evaluation was conducted? (Rog 2012)); and, (vi) dimensions.

Each intervention or program evaluation may be viewed as possessing a number of situated dimensions that describe the circumstances surrounding the issue; these could include physical, organisational, social, cultural, traditional, historical, or political dimensions. Rog’s contextual parameters basically set the scene within which an intervention occurs and may be subsequently evaluated. This model provides an organisational structure within which an intervention, such as that represented by the 2003 Australian Government reform package (Department of Education, Science and Training), may be examined through a program evaluation approach thus enabling increased understanding of the intervention and its longer-term influences and context outcomes.

Adaptation of Rog’s model to suit the current study

From the perspective of the evaluator, it was important to consider how Rog’s model could be adapted to unpack the context, which would then identify areas that required deeper investigation. Figure 3 presents the application of Rog’s contextual parameters, which guide the current study—moving from the macro context of international quality agenda through to the micro level context of the present study; these examples are provided to indicate the range and depth of literature that is required to support a rigorous investigation of context.



FIGURE 2: HIGHER EDUCATION REVIEW MILESTONES

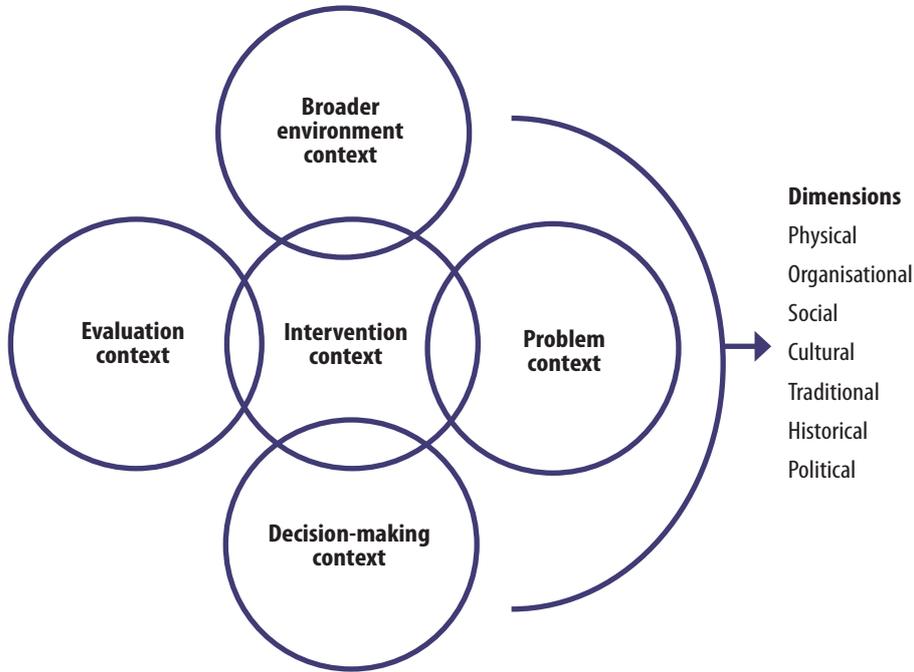
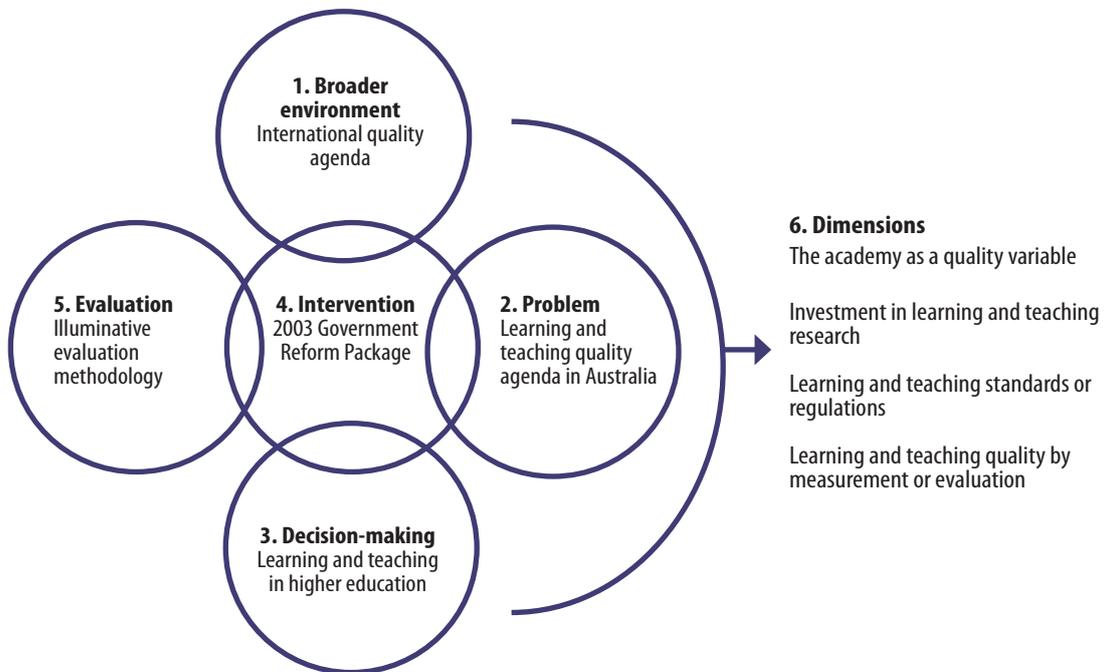


FIGURE 3: CONTEXTUAL PARAMETERS FOR THE CURRENT STUDY (ADAPTED FROM ROG 2012)





As part of the evaluation, a literature search was conducted through a number of search engines including Google Scholar, Google, Web of Knowledge, Australian Government and institutional websites to locate journal articles, books and government reports that would provide insight into the six specific areas identified as being context-sensitive to this intervention. What is critical in the fields of humanities, social sciences and education, is the fact that government-oriented research or agency-funded research may go under the radar (Matthews, et al. 2006). These types of 'grey' output and documents are not easy to locate through the usual search strategies and thus, often go undetected. Therefore, the inclusion of government and institutional websites became an important search strategy for this initial study.

Findings

The findings that emerged from this structured approach to unpack the context surrounding the intervention are presented in the six contextual parameters described below.

1. Broader environment: international quality agenda

This contextual parameter investigated the development of quality assurance in the international context; in particular, how it migrated from America to Europe and through to Australia. The origins and definition of quality assurance in higher education may be traced back to the late 1800s in America (Roades & Sporn 2002), where quality assurance emerged during a period of rapid growth in colleges and universities (Goldin & Fatz 1999). In America, quality assurance moved from business to government and through to education (Birnbaum 2000; Roades & Sporn 2002), where the model was modified to suit the new environment of higher education. This early practice of a self-regulatory, self-study model of quality assurance involved a five-step process: (i) the undertaking of an institutional self-review; (ii) a visit by an external panel of experts to the institution for a short period of review; (iii) the production of a report by the panel; (iv) the response provided by the institution; and (v) the final evaluation provided by the panel and made public. Even though Scriven (1991) criticised the limited nature of the modification as being superficial, this model remained standard practice in higher education in America for over a century. However, America's approach was situated in the local environment, managed by and focused on individual institutions (Roades & Sporn 2002). The adoption of this self-regulatory self-study model of quality assurance often required guidance from a national agency and in 1946, the American Society for Quality was established and is an early example of an agency that guided quality assurance practice. As an independent, non-profit organisation, it was comprised of a global

community of experts who offered quality guidance, best practice, training and policy advice to individual professionals, organisations, businesses, higher education institutions, and government departments.

After a century of practice in the United States of America, the trend towards quality assurance in higher education moved from America to Europe in the late 1980s and took the shape of a self-regulatory self-study model of quality assurance (Roades & Sporn 2002; Neave & van Vught 1994). Europe's adoption of the American model of quality assurance was directly influenced by three versions of isomorphism: coercive isomorphism; mimetic isomorphism; and normative isomorphism. Coercive isomorphism occurred when the American Government, which had adopted quality assurance from the business community, then coerced, or required, the higher education sector to adopt a modified version (Roades & Sporn 2002; Neave & van Vught 1994). As discussed below, coercion usually takes the form of funding or regulation. Mimetic isomorphism occurred when the European Union higher education sector, in the pursuit of market share against a strong competitor—America, mimicked, or copied the American quality assurance model in order to look similar to its competitors. Normative isomorphism occurred within the European academic community, where there is evidence to show that professional mechanisms (which include scholarly work, communities and networks, conferences, keynote speakers and increased interest in quality assurance), built a normative conduit for the adoption of ideas across the academic community and across international boundaries.

In 1993, the formation of the European Union through the Maastricht Treaty and the reconceptualisation of higher education through the Sorbonne Declaration in 1998 and the Bologna Declaration in 1999, led to change that went beyond national borders (Van der Wende 2000). In 1999, the European Union developed a strategic plan to reposition and reinforce higher education as a strong competitor (for international students) against America. The Bologna model of curriculum, as it became known, instigated the notion of borderless education, where students may start their undergraduate degree in one country and finish at another institution within the European Union (European Higher Education Area 2010). These declarations brought higher education institutions within the European Union to adopt a system of easily readable and comparable degrees that would facilitate student mobility. The adoption of the Bologna model brought into focus quality and how to ensure and assure that there are equivalences across each participating country, despite their diverse collection of institutions, policies and practices. After the introduction of the Bologna model, Europe continued to apply the American model of self-regulatory, self-study with the support and guidance of a national agency, the European Network



for Quality Assurance (2012). Established in 2000, the purpose of this network was to guide and disseminate information and good practice in quality assurance and to promote European cooperation in higher education.

In the 1990s, the United Kingdom was an early adopter of quality assurance in higher education within an environment of increasing accountability for public expenditure of funds and reduced funding (Harvey & Green 1993). The United Kingdom introduced a six-year cycle that applied an external audit standards model of quality assurance to the higher education sector (Harvey 2005) with a meta-national agency—the Quality Council (Neave 1994; van Vught 1991). This reform moved higher education in the United Kingdom towards a detailed, central regulation that was different from the European Union's model of guidance (Neave 1994; Neave & van Vught 1994). In a critical review of the history of quality evaluation in the United Kingdom, Harvey (2005) determined that quality assurance had become the primary policy for higher education.

In 1997, the United Kingdom established a second national agency, the Quality Assurance Agency (Harvey 2005). The United Kingdom received further criticism that its processes were burdensome, overly bureaucratic and more about compliance than improvement (Harvey 2005; Roades & Sporn 2002). The multiple layers of quality assurance contributed to the European perception that quality assurance in the United Kingdom had become unstoppable. However, it was this model that later directly informed the Australian quality assurance model emerging from the 2008 Review of Higher Education.

In summary, the model of quality assurance in higher education originated in America, moved to Europe and then ultimately to Australia. The early practice of a self-regulatory self-study model of quality assurance directly informed the learning and teaching quality agenda leading into the 2002 Government Review.

2. Problem context: learning and teaching quality agenda in Australia

This contextual parameter refers to the time when the government introduced two generations of performance evaluation and management. In 1983, a new Australian Labor Government introduced program evaluation to the public sector (Mackay 1998; 2004). Driven by the Department of Finance, the government was interested in tight control of public expenditure, conducting an annual budgetary process to inform its decision-making and improve the performance of the public sector. The purpose underpinning program evaluation has remained constant from 1987 to the present. The three main objectives were to: (i) provide fundamental information on program performance to assist decision-making; (ii) support the government's policy development and support and strengthen departments' internal

management (including staff learning); and, (iii) strengthen external reporting for accountability purposes (Mackay 1998; 2004). During this period of time, two different performance approaches to the management of program evaluation were adopted, as described below.

A critical analysis by the World Bank Operations Evaluation Department (Mackay 1988; 2004) analysed two generations of performance evaluation by the Australian Government. Basically, Australia went from a successful devolved system of performance evaluation, with an extensive library of evaluation portfolios, that supported and informed government decision-making, to a centrally-driven model that does not meet the needs of government. The earlier devolved model was strongly focused on the reporting of performance indicators, which reflected the change to a conservative government, and marked a change in the pursuit of public sector reform in Australia. As a result of this critical review, the World Bank study (Mackay 2004) identified that the more recent outputs–outcomes framework fell short of the government's goals and more explicitly— failed to inform their decision-making.

In addition to the findings by Mackay, there is a cautionary note found within the literature that identifies three obstacles to an intervention (Ryan 1999). These obstacles are: (i) the cost in terms of technical, personnel and time requirements; (ii) the complexity of the government environment at the time of the intervention; and, (iii) possible interference with the intervention by the political process. For example, in political terms, a government review of higher education in Australia is associated with a change of elected party in the national leadership (Department of Industry Innovation Science Research and Tertiary Education 2013). The 1988 Government Review was launched under the leadership of the Labor Party (Croucher, Marginson, Norton & Wells 2013); the 2002 Government Review was launched under the leadership of the Liberal-National Coalition; and the 2008 Government Review was launched by the Labor Party; these government reviews of higher education were designed to investigate the sector systematically, engage with relevant stakeholders and determine the need for future change.

There is a further cautionary note to be found within the literature. When a government places an interest in quality in higher education for public accountability purposes ahead of academic autonomy, it thereby deliberately undermines academic autonomy (Alderman 1996). There is no doubt that when expending public moneys, institutions should always be accountable. However, it is in the interest of all citizens to ensure that academic autonomy continues, as 'Academic autonomy is the lifeblood of higher education. It must not be stifled in the name of university politics or sacrificed on the altar of public accountability' (Alderman 1996, p. 192).



In summary, there are a number of cautionary notes within the literature that would suggest that any evaluation of a government intervention will require careful consideration of all the variables in play at the time. This reinforces the validity of applying Rog's context-sensitive approach to establish the foundation upon which a program evaluation can occur.

3. Decision-making context: learning and teaching in higher education in Australia

This contextual parameter explores the issues confronting the higher education sector in Australia by examining the context of each stakeholder at the turn of the twenty-first century. In 1995, the Australian Qualifications Framework was adopted within the national education policy setting. This framework offered a formal structure to benchmark educational standards and to guide the administration and granting of formal qualifications (Australian Qualifications Framework Advisory Board 2007; Ministerial Council for Tertiary Education and Employment 2011). The primary objectives included: (i) a national structure; (ii) flexible pathways; (iii) providing guidance to providers to meet requirements; (iv) flexibility to address diversity; (v) the encouragement of progress through all sectors; and (vi) the strengthening of vocational education and training, as well as recognition of all qualifications offered in Australia. The framework involved three education levels—primary, secondary/senior secondary, and tertiary; it was administered through three sectors: schooling; vocational education and training; and higher education.

Unlike other approaches to quality assurance models discussed earlier, the Australian Qualifications Framework was introduced without the support of any agency or regulator with oversight or responsibility for its implementation (Norton 2012). As a result, the centrally administered schooling and vocational education and training sectors adopted the framework, while the autonomous higher education institutions were able to choose whether to adopt the framework or not. As a result, the Australian higher education sector remained largely self-regulating throughout the rest of the 1990s, as well as the ensuing 2002 Government Review and 2003 Government Reform.

In 2000, Australia was at an economic crossroads. When the twentieth-century reliance on exploitation of natural resources for economic gain was faltering, the government realised the need for greater diversity in its sources of economic strength to reinforce its economic position in the world (Nelson 2003), and therefore looked to the higher education sector to play a more significant role in Australia's economic future. The Organisation for Economic Co-operation and Development (1996; 2004) found that there was a positive relationship between educational attainment

and a country's economic growth; this was strengthened when formal qualifications were a source of social capital for a country's citizenry (Sewell 1992). The Australian Government determined that a knowledge economy could support future economic stability; this would require an increase in the national focus on the higher education sector and especially on learning and teaching policy and practice.

Within this national push for a knowledge economy, public universities faced financial constraints when a new government funding model reduced funding from 90 per cent in 1981 to 75 per cent (Department of Education, Science and Training 2002a). This new model resulted in an overall reduction in real terms of 15 per cent to publicly funded institutions within the sector. The funding liability was shared between government (50 per cent) and students (25 per cent), with the remainder (25 per cent) being the responsibility of each institution. This reduced funding led to a growing interest in the efficiency of Australian universities and a perceived need for them to be more resourceful in their approaches to meet growing student demand (Abbott & Doucouliagos 2003). The student environment was directly impacted by widening participation, a more diverse student cohort and an increased student-to-staff ratio, at the same time as the cost of education was rising (Department of Education, Science and Training 2002a). There were more diverse pathways into university and this diversity itself brought challenges in terms of increasing needs of the student cohort (Hillman 2005). The diminished funding environment, with increased student cohort diversity, brought into question whether higher education providers could pursue quality maintenance, let alone the ability to improve quality with respect to agreed notions of excellence (Gyimah-Boadi 2003).

In summary, the Australian higher education sector faced government accountability, driven by financial accountability and a growing interest in standards and the quality agenda. Higher education providers were challenged by a reduced funding environment and needed to be more effective in their management and delivery. Meanwhile, the academy was coping with changing roles under a changing student environment. It was at this point that the 2002 Government Review of Higher Education in Australia was launched which led to the government's response in the form of the 2003 Reform Package.

4. Intervention context: 2003 Government Reform Package

This contextual parameter sets the central context of the 2003 Government Reform Package by exploring the cause surrounding policy initiatives, and the Government Review and Reform Package. In 2000, the Australian Government established the Higher Education Quality Assurance Framework (Department of Education, Science



and Training 2002a) to ensure a level of quality within qualifications offered by higher education providers that offered benchmarking opportunities across the sector and moved to invest in research. The framework had four main strategies including: (i) establish AUQA as a major national initiative to increase the accountability and quality for individual higher education providers; (ii) develop the National Protocols for Higher Education Approval Processes to provide protocols and guidelines to national and international institutions; (iii) commission the Quality Assurance and Accreditation in Australian Higher Education project to explore the nature and practice of accreditation, quality assurance and assessment in Australia and in a number of countries overseas; and, (iv) adopt a new model to address the strengths and weaknesses, and reposition quality assurance and accreditation in Australian higher education.

In 2002, AUQA (2007) was established and adopted as an externally regulated self-study model of quality assurance (Carroll 2003). This involved a five-step process: (i) the undertaking of an institutional self-review; (ii) a visit by an external panel of experts to the institution for a short period of review; (iii) the production of a report by the panel; (iv) the response provided by the institution; and, (v) the final evaluation provided by the panel and made public. There was a strong impetus to benchmark the quality of learning and teaching policy and practices for the higher education sector when AUQA was established in 2002. The purpose of this was to focus on an institutional level of review as a formal quality assurance process to evaluate learning and teaching, research, and the institution as a whole (Department of Education, Science and Training 2006). This national agency was directly linked to the historic model of quality assurance that had emerged from America, moved across to Europe, and ultimately to Australia, over a decade earlier.

In addition, the Australian Government, under the leadership of the Liberal-National Coalition, undertook a comprehensive review of the higher education sector to ensure that Australia's higher education institutions were best placed to contribute to the nation's future. Pressure was placed on the Australian higher education sector to meet the expectations of government, the community and students in such areas as accountability of funding, institutional reporting requirements, and a perceived need to meet student expectations. These pressures came from both the external challenges and internal pressures driving change in the higher education sector, with a complex meld of financial changes appearing to be most influential. As mentioned elsewhere, there were significant changes to the ways in which universities were funded at a national level (Department of Education, Science and Training 2002a). This meant there were a number

of national tensions driving change, and the government review of the higher education sector was, in effect, an intervention designed specifically to both identify and implement necessary change (Rog 2012).

In 2002, the scope of the Government Review extended across the higher education sector and outcomes were to impact directly on all higher education providers in Australia, with a strong emphasis on the reorientation of their focus towards learning and teaching (Ramsden 1991). Therefore, the focus of this government review represented a substantial policy shift by the Department of Education, Science and Training (2004) to focus attention on learning and teaching in higher education to meet the goals of the knowledge economy. As a result of the government review process, in 2003 the government published its response to the review, as a reform package entitled *Our Universities: Backing Australia's Future* (Nelson 2003). This 2003 Government Reform Package brought together a broad set of reforms underpinned by four key principles: (i) sustainability; (ii) quality; (iii) equity; and, (iv) diversity. There was an implementation timeframe for the reform package of 2006 to 2008, together with a planned evaluation of the reform in 2009 with a \$1.5 billion investment over four years. The focal point of this thesis is that of the 16 government reforms, 4 were of direct interest namely: promoting excellence in learning and teaching; enhancing collaboration; assuring quality; and, a new accountability framework.

In summary, the 2003 Government Reform Package implemented three national learning and teaching initiatives (LTPF, ALTC and AUQA) and these became the focus for the program evaluation.

5. Evaluation: illuminative evaluation methodology

This contextual parameter established the scope for the program evaluation under examination and clarified the aim and research questions guiding the evaluation. For example, the aim of the broader study was to document and theorise the consequences of the 2002 Government Review of Higher Education. This review led to the 2003 Government Reform Package on Learning and Teaching in higher education in Australia during 2002–2008 through an illustrative program evaluation perspective. From this central aim, the following three questions developed:

1. to what extent did the 2003 Government Reform Package focus specifically on learning and teaching in higher education?
2. what changes in learning and teaching may be identified in the period 2002 to 2008?
3. how did the 2003 Government Reform Package change the profile of the learning and teaching quality agenda in higher education in Australia?



At the point of intervention, Australian higher education was comprised of 38 autonomous public universities and 73 autonomous private providers (Department of Education, Employment and Workplace Relations 2008). Of the 38 public universities, 4 were dual-sector providers that offered both higher education, and vocational education and training qualifications; a further 17 were mixed-sector providers that offered higher education within a registered training organisation, and 1 was both a dual and mixed-sector provider (Moodie 2010). The overall student population consisted of approximately 1 million full-time-equivalent enrolments with an academic population of around 40 000 full-time equivalent staff (Department of Education, Science and Training 2006a).

The broader program evaluation represents an independent evaluation of the 2003 Government Reform Package and it adopted a methodology to conduct an evaluation of a large-scale illuminative program evaluation that is unique in its focus on learning and teaching. After analysis of a number of different evaluative approaches, illuminative evaluation was identified by the evaluator as an appropriate methodology (Parlett & Hamilton 1972; 1976; Patton 2002) to investigate both the intended and unintended outcomes of the three national learning and teaching initiatives. The full detail of the selection process, theory and methodology is covered in more depth in Alderman (2015).

In summary, the program evaluation applied illuminative evaluation as the evaluation methodology to elicit the intended and unintended consequences of the three national learning and teaching initiatives.

6. Dimensions

The final contextual parameter denotes how the dimensions relevant to this intervention were identified through an environmental scan of literature and practice of the circumstances surrounding the 2003 intervention, including the academy, investment in research, standards/regulations, measurement and evaluation. This process identified four sub-dimensions as outlined below.

The first sub-parameter explored the academy as a quality variable. As a result of this process, two variables were identified as having direct impact on the quality of learning and teaching in higher education: (i) the changing role of the academy (Coaldrake & Stedman 1999; 2013; McInnes 2000); and, (ii) casualisation of the academic workforce (Coates, Dobson, Goedegebuure & Meek 2009; Percy et al. 2008). As learning and teaching practice remains intrinsically linked to the quality agenda, there are direct consequences when government decision-making affects funding, focus and documentation.

The second sub-parameter surveyed the investment in learning and teaching research at the time. As mentioned earlier, there are a number of research outputs that are considered 'grey' literature, thus it was difficult to locate

research on learning and teaching research, resources to support educational research, and commissioned research into education sectors (McMeniman, Cumming, Wilson, Stenson & Sim 2000). Within each of these aspects, the higher education sector has some way to go to advance as a field of research and be as well supported as the vocational education and training sector in Australia.

The third sub-parameter investigated the learning and teaching standards or regulations in play at the time of the program evaluation. More specifically, it focused on the standards in learning and teaching through formal qualifications in learning and teaching (Australian Vice-Chancellors' Committee 1993); regulation through professional bodies; regulation through professional development (Chalmers 2007; 2010); regulation through pre-employment requirements; and, regulation through post-employment requirements. There were no national standards for academic staff engaged in a learning and teaching role in higher education in Australia (Department of Education, Science and Training 2002b; 2002c). While professional bodies, professional development, and employment requirements do offer opportunities for self-regulation, the extent of this practice remains under-researched.

The fourth and final sub-parameter reviewed evaluation of the quality of learning and teaching through measurement and government reporting within an outcomes and outputs framework; measurement through national student feedback surveys; measurement through a single voice of students; evaluation through benchmarking learning and teaching; evaluation through teaching excellence awards; evaluation through scholarship of teaching; and, evaluation through peer review (Kreber 2003; Productivity Commission 2010; Zammit et al. 2007). It deliberately applied the words 'measurement' and 'evaluation' to reflect the differences between the quantitative approaches taken and the qualitative approaches undertaken.

In summary, unpacking the dimensions pertaining to this intervention allowed the evaluator to determine other variables that may, or may not, directly impact on the intended outcomes. In particular, this sub-parameter identified a number of datasets that would assist the evaluator to establish the impact of the intervention through a systematic literature analysis.

Discussion

As demonstrated by the findings, the application of an adaptation of Rog's (2012) model of contextual parameters was found to be useful and informative on a number of levels. This process clarified a number of elements for the broader study, it: (i) sharpened the scope and focus on the 2003 intervention; (ii) confirmed the timeframe as valid; (iii) confirmed the value of the broader investigation (as no evidence was available to



indicate an evaluation of the intervention was planned or conducted); (iv) identified the main topics for the literature review, relevant to the broader topic; and, (v) identified illuminative evaluation as the appropriate evaluation method for the broader study.

First, and foremost, this process clarified the scope of the broader program evaluation. Originally, the evaluator had placed the 2002 Government Review of Australian Higher Education (Department of Education, Science and Training), and its associated documentation, as part of ‘the problem’. The application of Rog’s model highlighted the fact that the review was in fact part of the ‘decision-making context’, not ‘the problem’. Thus, as an outcome of this structured approach, the scope of the broader program evaluation was redefined to focus on the 2003 Australian Government Reform Package (Nelson 2003).

Second, this process identified situations where, regardless of the successful nature of the government policy implementations, some contextual parameters would dominate and interfere in the implementation. The intervention under examination in the broader program evaluation was located in the Australian higher education sector. Examination of previous higher education reviews and subsequent reform packages, identified that the policy implementation cycle may for disrupted should there be a change in federal government political leadership. This outcome confirmed that the change in political leadership was a significant variable for the broader study, as each year (2002 and 2008) signified a change in federal government political leadership, a subsequent review of Australian higher education (Bradley 2008; Department of Education, Science and Training 2002a) and confirmed the timeline for program evaluation.

Third, while the focus of the broader investigation was clarified, this process identified that the three national initiatives that focused on learning and teaching all had three-year funding cycles with an evaluation scheduled at the end of each cycle. However, there was no evidence of a similar schedule for the 2003 Australian Government Reform Package (Nelson 2003). There was evidence that an evaluation was planned, however, the change in government leadership initiated a subsequent review in 2008 (Bradley) that disrupted the planned evaluation in 2009. Therefore, this identified a gap in the literature and confirmed the value and purpose of the broader program evaluation.

Fourth, this approach provided a structured way to identify key articles and reports from the ‘grey’ literature and practice that were topical and pertaining to the specific time period. The process of identifying associated parameters enabled the literature to be separated by that which directly influenced the intervention, and that which was related to contextual

parameters. These dimensions became significant variables in the broader program evaluation.

Fifth, this initial study identified that a more holistic approach to evaluation was required in the broader program evaluation. Given the contextual parameters surrounding the intervention and the number of significant variables that would directly impact the intervention outcomes, the evaluator quickly realised that a more traditional evaluation methodology would not capture both the intended and unintended consequences of this intervention. It was through this process that the evaluator was able to identify illuminative evaluation by Parlett and Hamilton (1972; 1976) as the appropriate research methodology for the broader study.

Conclusion

In conclusion, this article defined the essence of the context-sensitive parameters and dimensions in play at the time of the intervention and program evaluation, through the application of an adaptation of Rog’s (2012) model of contextual parameters. This model offered the evaluator a structured approach to examine an intervention and this initial study provided a systematic way to clarify the scope, variables, timing, validity of study, and appropriate research methodology for the broader program evaluation. Given that the government implementation of an educational intervention under investigation did not follow the experimental research approach nor the double cycle of action approach, the initial study provided an in-depth understanding of the context-sensitive context. It is from this clear purpose that the broader evaluation was then conducted and overall, the investigation of the broader evaluation was strongly augmented by this initial approach. When governments or institutions implement policy to invoke educational change, program evaluation, or policy implementation analysis is achievable post-implementation. Rog’s (2012) model of context-sensitive parameters offers evaluators a structured approach to establish a solid foundation of understanding to achieve this outcome.

References

- Abbott, M & Doucouliagos, C 2003, ‘The efficiency of Australian universities: a data envelopment analysis’, *Economics of Education Review*, vol. 22, pp. 89–97.
- Alderman, L 2015, ‘Illuminative evaluation as a method applied to Australian Government policy borrowing and implementation in higher education’, *Evaluation Journal of Australasia*, vol. 15, no. 1, pp. 4–14.
- Alderman, L 2014, ‘From policy borrowing to implementation: an illuminative evaluation of learning and teaching in higher education in Australia (2002 to 2008)’, PhD thesis, Queensland University of Technology, Brisbane.



- Alderman, G 1996, 'Audit, assessment and academic autonomy', *Higher Education Quarterly*, vol. 50, no. 3, pp. 178–192, DOI: 10.1111/j.1468-2273.1996.tb01700.x.
- Australian Qualifications Framework Advisory Board 2007, *Australian Qualifications Framework: implementation handbook*, Australian Qualifications Framework Advisory Board to Ministerial Council for Education, Employment, Training and Youth Affairs, Carlton South, Victoria.
- Australian Universities Quality Agency 2007, *Mission, objectives, vision and values*.
- Australian Vice-Chancellors' Committee 1993, *Guidelines for effective university teaching*, Australian Vice-Chancellors' Committee, Canberra.
- Birnbaum, R 2000, 'The life cycle of academic management fads', *The Journal of Higher Education*, vol. 71, no. 1, pp. 1–16.
- Bradley, D 2008, *Future directions for tertiary education*, Department of Education, Employment and Workplace Relations, Canberra.
- Bugelski, BR 1971, *The psychology of learning, applied to teaching*, 2nd edn, The Bobbs-Merrill Company Inc, Indianapolis, Indiana.
- Carroll, M 2003, 'Does auditing he against standards encourage masterpieces or paint-by-numbers?', *Assessment and Evaluation in Higher Education*, vol. 28, no. 3, pp. 297–307, DOI: 10.1080/0260293032000059649.
- Chalmers, D 2010, *National teaching quality indicators project – final report*, The University of Western Australia, Perth.
- Chalmers, D 2007, *A review of Australian and international quality systems and indicators of learning and teaching*, The Carrick Insitute for Learning and Teaching in Higher Education Ltd, Strawberry Hills, NSW.
- Coaldrake P & Stedman L 2013, *Raising the stakes*, University of Queensland Press, St Lucia, Queensland.
- Coaldrake, P & Stedman, L 1999, *Academic work in the twenty-first century: changing roles and policies*, Department of Education, Training and Youth Affairs, Canberra.
- Coates, H, Dobson, I, Edwards, D, Friedman, T, Geodegebuure, L & Meek, L 2009, *The attractiveness of the Australian academic profession: a comparative analysis*, LH Martin Institute and the Australian Council for Educational Research, Melbourne.
- Croucher, G, Marginson, S, Norton, A & Wells, J 2013, *The Dawkins revolution 25 years on*, Melbourne University Press Melbourne.
- Department of Education, Employment and Workplace Relations 2008, *Students: 2008 summary of higher education statistics*, Commonwealth of Australia, Canberra.
- Department of Education, Science and Training 2006, *Handbook for pilot quality audits of higher education providers*, Commonwealth of Australia, Canberra.
- Department of Education, Science and Training, 2004a, *Learning and teaching performance fund 2004*, viewed 20 March 2006, http://pandora.nla.gov.au/pan/42033/20040519-0000/www.dest.gov.au/highered/learning_teaching/p_fund_default.htm.
- Department of Education, Science and Training 2003, *Our universities backing Australia's future*, Commonwealth of Australia, Canberra.
- Department of Education, Science and Training 2002a, *Higher Education at the crossroads*, Commonwealth of Australia, Canberra.
- Department of Education, Science and Training 2002b, *Raising the standards: A proposal for the development of an ICT competency framework for teachers*, Commonwealth of Australia, Canberra.
- Department of Education, Science and Training 2002c, *Raising the standards: Appendices to the proposal for the development of ICT competency framework for teachers*, Commonwealth of Australia, Canberra.
- Department of Industry, Innovation, Science, Research and Tertiary Education 2013, *Federal, state and territory election dates from 1946 to the present*, viewed 7 April 2013, http://www.aec.gov.au/Elections/Australian_Electoral_History/Federal_State_and_Territory_elections_dates_1946_Present.htm.
- European Higher Education Area 2010, *Budapest-Vienna Declaration on the European Higher Education Area*, viewed 8 January 2012, http://www.ond.vlaanderen.be/hogeronderwijs/bologna/2010_conference/documents/Budapest-Vienna_Declaration.pdf.
- European Network for Quality Assurance (ENQA) Secretariat 2012, *Welcome to the website of ENQA*, viewed 15 January 2012, <http://www.enqa.eu>.
- Gall, JP, Gall, MD & Borg, WR 2005, *Applying educational research*, Pearson Education Inc, Boston.
- Goldin, C & Fatz, LF 1999, 'The shaping of higher education: the formative years in the United States, 1890 to 1940', *Journal of Economic Perspectives*, vol. 13, no. 1, pp. 37–62.
- Gyimah-Boadi, E 2003, 'Reflections on education and democracy', *Ghana Centre for Economic Development*, vol. 6, no. 2, pp. 1–5.
- Harvey, L 2005, 'A history and critique of quality evaluation in the UK', *Quality Assurance in Education*, vol. 13, no. 4, pp. 263–276, DOI: 10.1108/09684880510700608.
- Harvey, L & Green, D 1993, 'Defining quality', *Assessment and Evaluation in Higher Education*, vol. 18, no. 1, pp. 9–26. DOI: 10.1080/0260293930180102.
- Heafner, R 1932, *The typewriter in the primary and intermediate grades*, The Macmillan Company, New York.
- Kemmis, S & McTaggart, R 1988, *The Action Research Planner*. Deakin University, Melbourne.
- Kemmis, S, McTaggart, R & Retallick, J (eds.) 2004, *The action research planner*, 2nd edn revised, Aga Khan University, Institute for Educational Development, Karachi, Pakistan.
- Kogan, M 2005, 'The Implementation Game', in A Gornitzka, M Kogan and A Amaral (eds), *Reform and change in Higher Education: analyzing policy implementation*, Springer, Dordrecht, Netherlands, pp. 57–65.
- Kreber, C 2003, 'The relationship between students' course perception and their approaches to studying in undergraduate science courses: a Canadian experience', *Higher Education Research & Development*, vol. 22, no. 1, pp. 57–75, DOI: 10.1080/0729436032000058623.
- Lingard, B 2010, 'Policy borrowing, policy learning: testing times in Australian schooling', *Critical Studies in Education*, vol. 51, no. 2, pp. 129–147.
- Lingard, B & Garrick, B 1997, 'Producing and practising social justice policy in education: a policy trajectory study from



- Queensland, Australia', *International Studies in Sociology of Education*, vol. 7, no. 2, pp. 157–179.
- Mackay, K 2004, *Two generations of performance evaluation and management system in Australia*, The World Bank Operations Evaluation Department, Washington DC.
- Mackay, K 1998, *The development of Australia's evaluation system*, World Bank Operations Evaluations Department, Washington DC.
- Maggetti, M, Gilardi, F & Radaelli, C 2012, *Research design in the social sciences*, SAGE Publications, Zurich.
- Matthews, D, Biglia, B, Henadeera, K, Desvignes-Hicks, J, Faletic, R & Wenholz, O 2006, *A bibliometric analysis of Australia's international research collaboration in science and technology: analytical methods and initial findings*, Forum for European–Australian Science and Technology Cooperation (FEAST), Canberra.
- McInnes, C 2000, 'Changing academic work roles: the everyday realities challenging quality in teaching', *Quality in Higher Education*, vol. 6, no. 2, pp. 144–152, DOI: 10.1080/713692738.
- McMeniman, M, Cumming, J, Wilson, J, Stevenson, J & Sim, C 2000, 'Teacher knowledge in action', in Higher Education Division Department of Education Training and Youth Affairs (ed.), *The impact of educational research: research evaluation programme*, pp. 375–550, Commonwealth of Australia, Canberra.
- Ministerial Council for Tertiary Education and Employment (MCTEE) 2011, *Australian Qualifications Framework: AQF levels and qualification type descriptors*, Ministerial Council for Tertiary Education and Employment (MCTEE), Carlton South, Victoria.
- Moodie, G 2010, *Mixed-sector tertiary education: implications for self-accrediting and other higher education institutions*, National Centre for Vocational Education Research (NCVER), Canberra.
- Neave, G & van Vught, F 1994, *Government and higher education relationships across three continents: the winds of change*, Permagon Press and IAU Press, Oxford and Paris.
- Neave, G 1994, 'The politics of quality: developments in higher education in Western Europe 1992–1994', *European Journal of Education*, vol. 29, no. 2, pp. 115–134.
- Nelson, B 2003, *Our universities: backing Australia's future*, Department of Education, Science and Training, Canberra.
- Norton, A 2012, *Mapping Australian higher education*, Grattan Institute, Melbourne.
- Organisation for Economic Co-operation and Development (OECD) 2004, *Policy brief: lifelong learning*, OECD Publishing, Paris.
- Organisation for Economic Co-operation and Development (OECD) 1996, *Lifelong learning for all*, OECD Publishing, Paris.
- Owen, JM & Rogers, P 1999, *Program evaluation: forms and approaches*, Allen & Unwin, St Leonards, NSW.
- Parlett, M & Hamilton, D 1976, 'Evaluation as illumination', in D Tawney (ed.), *Curriculum evaluation today: trends and implications*, Macmillan Education, London.
- Parlett, M & Hamilton, D 1972, *Evaluation as illumination: a new approach to the study of innovative programs*, Edinburgh University, Centre for Research in the Educational Sciences, London.
- Patton, MQ 2002, *Qualitative research & evaluation methods*, 3rd edn, Sage Publications, Thousand Oaks, California, London, Delhi.
- Percy, A, Scoufis, M, Parry, S, Goody, A, Hicks, M, Macdonald, I, et al. 2008, *The RED report: recognition, enhancement, development*, Australian Learning and Teaching Council, Sydney.
- Ponzi, LJ & Koenig, M 2002, 'Knowledge management: another management fad?', *Information Research*, vol. 8, no. 1, pp. 1–9.
- Productivity Commission 2010, *Vocational education and training workforce, productivity commission draft report*, viewed 8 January 2012, <http://www.pc.gov.au>.
- Ramsden, P 1991, 'A performance indicator of teaching quality in higher education: the course experience questionnaire', *Studies in Higher Education*, vol. 16, no. 2, pp. 129–150, DOI: 10.1080/03075079112331382944.
- Roades, G & Sporn, B 2002, 'Quality assurance in Europe and the U.S.: professional and political economic framing of higher education policy', *Higher Education*, vol. 43, no. 3, pp. 355–390, DOI: 10.1023/A:1014659908601.
- Rog, DJ 2012, 'An introduction to context and its role in evaluation practice', in DJ Rog, JL Fitzpatrick & RF Conner (eds.), *Context: a framework for its influence on evaluation practice*, Jossey-Bass, Vancouver, Canada.
- Ryan, N 1999, 'Rationality and implementation analysis', *Journal of Management History (Archive)*, vol. 5, no. 1, pp. 36–52, DOI: 10.1108/13552529910249832.
- Scriven, M 1991, *Evaluation thesaurus*, 4th edn, Sage, Newbury Park, California.
- Sewell, WH 1992, 'A theory of structure: duality, agency and transformation', *American Journal of Sociology*, vol. 98, no. 1, pp. 1–29.
- Van der Wende, MC 2000, 'The Bologna Declaration: enhancing the transparency and competitiveness of European higher education', *Higher Education in Europe*, vol. 25, no. 3, pp. 305–310, DOI: 10.1080/03797720020015890.
- Van Vught, F 1991, 'Higher education quality assessment in Europe: the next step', paper presented to the 39th bi-annual conference of the CRE, Utrecht, Netherlands.
- Weiss, CH 1983, 'The stakeholder approach to evaluation: origins and promise', *New Directions for Program Evaluation*, no. 17, pp. 3–14.
- Wood, BD & Freeman, FN 1932, *An experimental study of the educational influences of the typewriter in the elementary school classroom*, The Macmillan Company, New York.
- Zammit, K, Sinclair, C, Cole, B, Singh, M, Costley, D, Brown a'Court, L, et al. 2007, *Teaching and leading for quality Australian schools*, Teaching Australia, Australian Institute for Teaching and School Leadership Ltd, Sydney.



RALPH RENGER

System evaluation theory (SET): a practical framework for evaluators to meet the challenges of system evaluation

This article presents a guiding framework for evaluators to conduct systems evaluation: system evaluation theory (SET). The article firstly defines 'systems', 'system theory', and 'system thinking'; it then discusses the confusion in the evaluation literature surrounding the use of these terms (noting their application is primarily in evaluating programs, not evaluating modern day systems). Three guiding principles for evaluating modern day systems are then presented demonstrating: a) their relationship to systems theory; and b) how they assist the evaluator in accomplishing an evaluation.

Program evaluation literature is currently seeing an influx of articles devoted to system thinking and system theory (Von Bertalanffy 1968; Williams & Hummelbrunner 2010; Adams et al. 2014; Williams 2015). The potential of system thinking to provide meaningful information upon which to make programmatic decisions is tremendous. This is primarily because system thinking is postulated to lead to a better representation of the truth (Rogers 2008; Williams & Hummelbrunner 2010); thus addressing many of the criticisms levied against program and policy evaluation approaches that typically view processes as linear, in isolation, and removed from context (Morell 2005; Williams & Hummelbrunner 2010).

However, trying to accurately depict and consider context in a program evaluation adds complexity, and with it, consternation about its use in the evaluation community (Renger et al. 2012). Some evaluators refer to the challenge of considering context in the evaluation as 'wicked, messy and horribly tangled' (Williams & Hummelbrunner 2010, p.1). System thinking and system theory are offered as a solution to untangling this mess; these terms will now be defined. Ericson (2011) defines a *system* as:

An integrated composite of components that provide function and capability to satisfy a stated need or objective. A system is a holistic unit that is greater than the sum of its parts. It has structure, function, behavior, characteristics, and interconnectivity. Modern day systems are typically composed of people, products, and environments that together generate complexity and capability (p. 402).



Ralph Renger is a Professor and Director of the Division of Evaluation at the Center for Rural Health, University of North Dakota.
Email: ralph.renger@med.und.edu



A system consists of parts, subsystems, relationships, feedback mechanisms, attributes, and inputs, that interact in the surrounding environment and work toward a common goal (Buckley 1967; 1998; Burns 2007; Cohen & Kibel 1993; Meadows & Wright 2008; Williams & Hummelbrunner 2010). Banathy (1992) defines *system theory* as seeking:

A likeness of concepts, principles and laws that exist in the various realms of experience and cuts across all the domains of scientific disciplines. The arrangement of these general concepts, principles and laws constitutes a General Theory of Systems: an exposition applying to all systems. This power of integration across all disciplines is needed so complex systems can be understood (p. 10).

System theorists use ‘holistic approaches to analysis that focuses on the way a system’s constituent parts interrelate and how systems work over time and within the context of larger systems’ (Miller-Williams & Kritsonis 2009, p. 2). Further, system theorists apply system thinking to identify robust system principles to establish factors applicable to all system types (Von Bertalanffy 1968). Checkland (1999) defines *system thinking* as the process of thinking using systems ideas, rather than using system theory precisely (p. 10).

To better understand the potential of systems thinking and system theory in advancing the field of evaluation, the Deutsche Gesellschaft für Internationale Zusammenarbeit, GmbH (GIZ) hosted an international conference specifically devoted to the topic of *system evaluation*. A handful of the world’s leading experts were invited to the conference to share their perspectives regarding the current state and potential of system evaluation (Rogers 2011). The conference conclusions were mixed. There was optimism about the potential of system thinking and system theory. However, there was concern about the lack of a shared understanding on some of the most basic concepts, for example, the definition of a system evaluation, the difference between system thinking and system evaluation, and so forth (Lanzendörfer et al. 2011). This led to frustration and confusion among some evaluators (Williams & Hummelbrunner 2010), leading others to question the readiness of our discipline to conduct high quality system evaluations (Renger et al. 2012).

I was invited to the conference as a speaker and subsequently attended numerous seminars on the conference theme by leading experts in our field. I believe one important step in achieving clarity is to understand the current evaluation discussion remains centered on how systems thinking and systems theory can improve program evaluations. For example, Williams and Hummelbrunner (2010) provide 19 mixed-methods grounded in system thinking and system theory for augmenting *program* evaluation. To achieve clarity, it is

necessary for evaluators to articulate whether they are advocating for adopting system thinking to augment *program* evaluation (a topic of interest since the early 1990s) (Cohen & Kibel 1993), or whether they are referring to using system thinking and system theory to assist in evaluating *systems*; the latter receiving virtually no attention in the evaluation literature.

The focus of this article is the use of systems thinking and systems theory in evaluating systems not programs.

McDavid, Huse and Hawthorn (2013) define a *program* as something smaller in scale than a system and for which interaction with other subsystems is not necessary for its functioning. Further, although a program can benefit from feedback, it does not require feedback to successfully achieve its intended goals (Borich & Jemelka 1982). For example, a continuing education program can repeatedly deliver the same didactic information, without feedback or interaction with other subsystems. The feedback distinction between a system and a program can be a source of confusion. Evaluations often provide feedback as part of a program evaluation; this is not the same as the built-in feedback mechanism of a system used to maximize its efficiency and effectiveness.

As Weiss (1995) famously noted, there is nothing so practical as a good theory. There are many theories to guide the evaluator in completing a program evaluation (Alkin & Christie 2004; Mark, Henry & Julnes 2000; Fetterman 1994). However, there is no system evaluation theory, that is, there are no guidelines for how to conduct a system evaluation. A system evaluation theory is needed to provide evaluators with a blueprint to complete evaluations of systems, which, because of their scale, interconnectedness, and interaction with the environment, are likely to be more ‘messy’ and complicated than program evaluations. Our discipline is in desperate need of a system evaluation theory. Therefore, the purpose of this article is to offer an evaluation theory for conducting system evaluation (i.e. SET).

SET is grounded in system theory; the tenets of system theory are well established and proven, thus giving SET a solid theoretical basis. SET focuses on describing how proven system theory principles can guide the evaluator in accomplishing an evaluation of a modern day system, defined as consisting of human actors interacting with each other and/or technology (Buckley 1998; Ericson 2011).

SET consists of three guiding principles derived directly from system theory and the consideration of evaluation theories based in use, value, and methods (Christie & Alkin 2008); they are referred to as guiding principles so as to: a) differentiate them from system theory principles; and, b) reinforce their purpose to guide the practitioner in the ‘how to’ of conducting a system evaluation.

The application of the SET guiding principles was tested in the context of evaluating emergency response systems to time critical events (e.g. cardiac arrest, stroke)



in five US states. Each SET guiding principle and the corresponding system theory principle(s) from which it is derived is described below. The system theory principles are italicized so the reader can seamlessly follow how system theory underpins SET.

It is worth noting the three most commonly cited system theory principles in the program evaluation literature are interrelationships, perspectives, and boundaries (Williams & Hummelbrunner 2010). The set of system theory principles upon which SET is based are more comprehensive because the focus of SET is on evaluating systems, not identifying methods for improving program evaluations.

The goals of the article are to: a) lay the foundation for evaluators to embrace system evaluation when appropriate; b) provide a needed blueprint for proceeding with the evaluation of systems; and, c) serve as a springboard for advancing system evaluation theory.

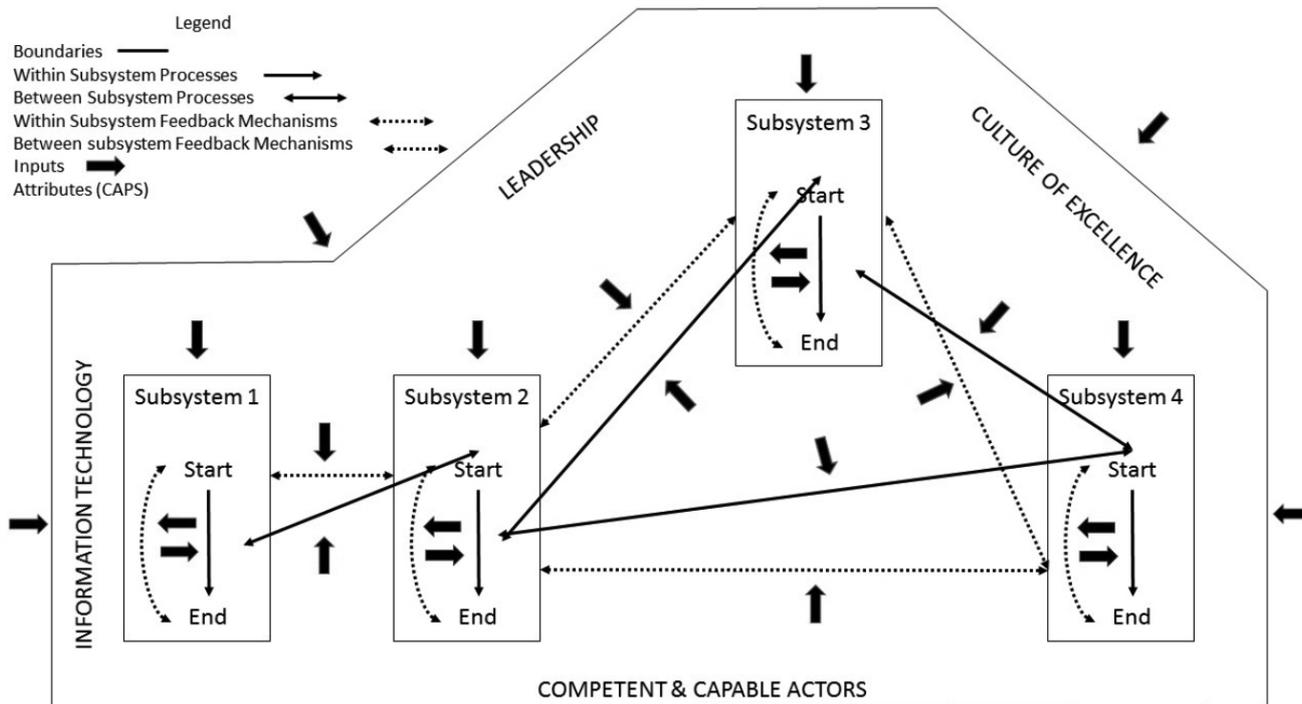
The three guiding principles of systems evaluation theory

SET Guiding Principle 1

It is necessary to define the system before evaluating efficiency and effectiveness. This includes defining the system boundaries, subsystems, processes, relationships, feedback mechanisms, attributes, inputs and common goal(s).

A hypothetical system depicting key system elements that need to be defined is shown in Figure 1. Within SET, every line, arrow, box, and text in Figure 1 designates a specific system element needing consideration in planning and conducting a system evaluation.

FIGURE 1. DEFINING KEY SYSTEM ELEMENTS FOR THE SYSTEM EVALUATION PLAN





Guiding Subprinciple 1a

Define system boundaries

The system boundaries establish the start and endpoint of what is to be evaluated. In Figure 1, the boundary is depicted as a solid line surrounding the system. The starting point is defined by the *stimulus* (Gibson 1960) that begins the system working toward its common goal. By extrapolation then, the end point is where the system stops working toward the goal, regardless of whether it is achieved. Being clear about the boundaries is critical for several reasons—first, it will affect the scale and the cost of the evaluation; the broader the boundaries, the greater the scale and cost. Second, and perhaps more importantly, is understanding that the boundaries establish whose *perspectives* (Williams & Hummelbrunner 2010) need to be included in the evaluation; ensuring the proper perspectives are included is essential to building the credibility of the evaluation and to ensure the evaluation results are used (Patton 2008).

Using system leaders to define boundaries is preferred. Leaders will be able to leverage the resources and bridge subsystems to secure cooperation to complete the evaluation (Patton 2008). Therefore, involving leaders from the onset is critical to gaining buy-in (Mohan 2014). Without strong leadership, conducting a successful system evaluation is unlikely (Patton 2008).

Where a system starts and ends is open to interpretation. For example, in emergency response, some experts argue the response begins with the emergency dispatch. However, the case can be made that the stimulus initiating the response is the bystander call to the dispatch. Therefore, it can be argued public education surrounding how to make an emergency call defines the initial system boundary. For these reasons it is important to validate the boundaries by asking whether there are any upstream or downstream subsystems that should be included prior to, or following the start and end boundary, respectively. That is, leaders should be challenged as to whether the boundaries are correct by extending the boundaries in each direction, and assessing whether the new boundaries are appropriate or are an overextension of the client's desired scope.

Guiding Subprinciple 1b

Define subsystems and subsystem boundaries

One defining feature of systems is *complexity* (Stame 2004). As the number of working parts in a system increases, so does the complexity. Complexity may also be added through the presence of embedded *subsystems*

(Andrews, Odum & Odum 1971), each with their own working parts. For example, the emergency response system consists of law enforcement, emergency medical services, and hospital subsystems (Eisenberg 2013). Therefore, once the system boundaries are established, it is necessary to define existing subsystems and their respective boundaries (Andrews, Odum & Odum 1971). The subsystem boundaries in Figure 1 are shown as a solid line surrounding each subsystem and are defined by the start and end point of the subsystem processes (see Subprinciple 1c).

The importance of defining the subsystems cannot be overstated. Other work devoted to systems evaluation emphasizes defining the system boundaries (Williams & Hummelbrunner 2010), but fails to recognize the importance of defining the respective subsystem boundaries. Once again, *perspectives* becomes an important consideration (Williams & Hummelbrunner 2010). By virtue of their organizational position, system leaders will be able to identify key subsystems. However, to understand subsystem boundaries, it is important to identify and engage subsystem leaders. Establishing relationships with subsystem leaders during the initial stages of defining the system is essential to securing cooperation needed by the evaluator to: a) develop a feasible evaluation plan that is meaningful to the user; and, b) have the evaluation results used (Patton 2008).

Guiding Subprinciple 1c

Define within subsystem processes

Having defined the subsystems and their respective boundaries, it is then necessary to define how each subsystem functions, that is, the operational steps. The operational steps of each subsystem are denoted in Figure 1 as 'start' and 'end'. Start and end are connected by a solid arrow indicating there are likely many operational steps and decision points between the start and end boundary.

Explicitly stated subsystem processes are necessary in evaluating the quality of *system feedback mechanisms* (Flynn, Schroeder & Sakakibara 1994; Jeliakova & Westerheijden 2002). Understanding what to do, as defined by the system process, is a necessary prerequisite for efficient and effective system functioning. Feedback mechanisms can then be aligned to ensure they provide specific information about corrective actions to improve the efficiency of subsystem processes (see guiding Subprinciple 1e).

System theory suggests soft systems methods, such as a process model (Checkland & Poulter 2010), may be useful in defining subsystem functioning. The evaluation literature on continuous quality improvement is useful in providing evaluators an array of methods for defining



subsystem processes (Abdulmalek & Raigopal 2007; Radnor et al. 2006; Holweg 2007; Øvretveit & Gustafson 2002). Standard operating procedures (SOPs) or other readily available source documentation are useful in detailing processes (Renger 2011). Using available SOPs are advantageous because they are endorsed and used by system actors. Describing subsystem processes using system actor terminology will facilitate evaluation buy-in (Patton 2008). Further, using existing source documentation also honors the feasibility evaluation standard, as it is a cost-effective method for defining the subsystem processes (Sanders 1994).

When passive approaches for defining subsystem processes (such as the use of source documentation are not possible), then active data collection strategies may be required. The importance of established relationships with system and subsystem leaders becomes salient at this stage, for it is they who are critical in identifying and securing the cooperation of other needed perspectives to help the evaluator define subsystem processes. Engaging subsystem subject matter experts (SMEs) directly, using approaches like process flow mapping, have proven effective in explicating and validating each subsystem operational step (Renger et al., in press b). Compared to passive approaches (such as the examination of existing standard operating procedures), process flow mapping is more expensive because it requires engaging SMEs throughout the system. However, the return on investment for process flow mapping is high, because SMEs have direct input into the evaluation plan, thus improving their buy-in and likelihood of using the results (Becker, Renger & McPherson 2015).

Guiding Subprinciple 1d

Define between subsystem processes (i.e. relationships and communication)

Subsystems are *interconnected* and therefore, need to be evaluated for possible cascading failures, that is—catastrophic domino effects (Buzna, Helbing & Peters 2008; Parsons 1961). To do this, it is essential to define how subsystems relate or communicate with each other. The initial communication of one subsystem to another (which triggers the second subsystem into action) can occur anywhere within the first subsystem process. Once communication is initiated, it is often bidirectional; for this reason, a solid two-way arrow is drawn from the heart of the upstream subsystem processes to the start of the downstream subsystem process (see Figure 1).

The relationship or communication processes between subsystems can be defined using the same methods as those for defining within subsystem

processes. Once the communication protocols between subsystems are defined, it is important to bring together subsystem leaders and SMEs to be sure there is a shared understanding of each other's communication protocols. Evaluators may find the literature on evaluating emergency preparedness initiatives especially helpful in designing approaches to evaluate between subsystem communication processes. For example, the Homeland Security Exercise Evaluation Program describes a building block approach for developing and testing communication protocols (Department of Homeland Security 2013).

Guiding Subprinciple 1e

Define system feedback mechanisms

Systems must continually monitor the environment and make corrective actions in order to survive and grow (Brandon 2014). Therefore, the evaluation must determine whether needed system feedback mechanisms (Flynn, Schroeder & Sakakibara 1994) exist and if so, how well they are functioning.¹ Two types of feedback mechanisms need to be evaluated—those within subsystems and those between subsystems; these mechanisms are depicted in Figure 1 as dotted dual arrows.

When evaluating the quality of existing system feedback mechanisms, it is important to distinguish between quality improvement and quality assurance (Fainter 1991). Since the system is constantly striving to improve efficiency, the evaluation must focus on ensuring the subsystems are engaged in continuous quality improvement as opposed to quality assurance. Quality assurance focuses on adherence to an established standard rather than seeking ways to exceed the standard (Fainter 1991). Since a system must continually adjust to its environment, quality improvement approaches are more germane to system evaluation.

Guiding Subprinciple 1f

Define system attributes

Attributes are system characteristics impacting its efficient and effective functioning (Juergens 1977). SET defines four core attributes that evaluators must consider when evaluating modern day systems. The core attributes are derived from the simultaneous consideration of system theory and a review of human systems literature (Buckley 1967, 1998; Eisenberg 2013). The four core attributes are the system's need for competent and capable system operators; committed



leadership; necessary IT infrastructure; and, a healthy organizational culture (Eisenberg, 2013). Figure 1 shows how these four attributes (capitalized in Figure 1) affect all aspects of the system, including within and between subsystem processes, feedback mechanisms, and inputs.

SET core attributes

1. Competent and capable system operators

Many systems depend on some level of human interaction to operate. Therefore, it is essential that the evaluation validate whether individuals interacting with the system have the necessary competency and capabilities to efficiently operate within the system; these individuals are referred to as system actors.

Systems can potentially consist of hundreds of actors across numerous subsystems. The broader the system boundaries, the less likely the resources will exist to evaluate all subsystems—and the more difficult it is to abide to the feasibility evaluation standard (Sanders 1994). Therefore, evaluators will need to leverage existing system infrastructure to evaluate actor competencies and capabilities. For example, the evaluation may need to exploit existing credentialing and certification processes to ensure system actors are competent, capable, and operating efficiently.

As the system boundaries expand ways to build evaluation capacity and empower systems actors to self-monitor, competencies and capabilities may be necessary (Boyle & Lemaire 1999). Developing evaluation tools, such as those grounded in behavioral anchors, can assist system actors in ensuring they maintain the highest competency and capability levels by providing real-time evaluation (Arvey et al. 1973; Kazdin 1979; Schwab, Heneman III & DeCotiis 1975).

2. Committed leadership

Leadership is another attribute critical to system efficiency and effectiveness. There is abundant psychological, management, sport, and evaluation literature noting the importance of leadership to the success of systems (Cohen & Tichy 2003; Fullan 2004; Reed 2006). Leaders can: a) provide necessary direction; b) motivate; c) provide leverage; d) delegate; and, e) use positional or political power to ensure the system receives the inputs necessary for survival and success (Adair 2009; Heifetz & Laurie 1997). Evaluators must determine whether the leadership exists at the different system levels for efficient and effective system functioning (Curphy, Hogan & Hogan 1994).

3. Necessary information technology infrastructure

One core attribute of many modern day systems is the dependence on IT (Bakos & Treacy 1986). The reality is the flow of information and feedback mechanisms in

many societal systems is technology dependent. This can be as simple as email correspondence or as complex as databases programmed to provide real-time data exchange. The evaluator must be aware of the potential *cascading effect* caused by system interoperability issues (Baker & Manoj 2007; Buzna, Helbing & Peters 2008), and may need to actively seek out IT expertise to understand current system information flow, identify potential bottlenecks, and propose corrective actions.

4. Organizational culture

Defining culture is difficult because the notion can often be unclear and misunderstood (Gershon, et al. 2004; Scott, Mannion & Marshall 2003). Compounding this difficulty is the nature of culture itself—it is a complicated phenomenon produced and consumed by multiple entities (Scott et al. 2003). Schein (1990) defines organizational culture as the shared norms, values, and operating assumptions of an organization that ultimately guide its members' internal and external behavior. Organizational culture is comprised of three components: 1) espoused values; 2) observable artifacts; and, 3) basic underlying assumptions (Schein 1990). Identifying and understanding organizational constructs is important because it helps to move beyond the superficial *how* and *what* behind organization initiatives and gain deeper insight to the *why*. Moreover, culture is a good indicator of the sustainability, adaptability, and overarching success of organizations.

Organizational and anthropological theories are useful in evaluating whether the culture is supportive of meeting the system goals. For example, cultural relativism recognizes it is critical to evaluate a culture against its own standards (Donnelly 1984). The challenge for evaluators is defining the appropriate cultural standards. A multi-method approach is appropriate to capture the more robust, accurate, and useful insights (Scott et al. 2003); for instance, the anthropological theory of cultural relativism is useful in evaluating whether a culture is supportive of meeting the system goals, by evaluating a culture against its own standards (Donnelly 1984).

Reviewing policies and procedures (e.g. a mission statement) may provide some insight into the organizational culture. However, this may not lead to an accurate representation of the truth—what is written and what is done may be quite different. The evaluator must continually monitor all interactions with leaders and system actors for clues about the state of the system culture through multiple communication modalities and process interactions (Scott et al. 2003).

Evaluating culture is challenging, however, it is reasonable to assume it is intricately intertwined with the system leadership (Andrews, 1985). Where leadership is evident, a healthy culture is more likely to follow (Kotter, 2011).

**Guiding Subprinciple 1g**

Define system inputs

Inputs include any resources or assets needed to ensure system actors, processes, feedback mechanisms, and so forth, operate efficiently. Inputs are well understood by theory-driven evaluators, as they are an essential element of the logic model (Deyo, Julian & Jones 1995). As shown in Figure 1 through the numerous solid blocked arrows, inputs are needed at multiple points for efficient and effective system functioning to support system attributes, within and between system processes and feedback mechanisms.

SET recommends simultaneously defining system inputs as boundaries and processes are defined. This is because each step in the system process will require *energy* or some level of resources to complete. It is easier for SMEs to specify the needed resources as each step in the process model is defined because it is placed in context during process flow mapping.

Guiding Subprinciple 1h

Define the common systems goals

SET defines two types of system goals: those related to efficiency and those related to effectiveness. Defining efficiency goals is similar to the theory-driven evaluation concept of implementation theory, where activities need to be bound together by an underlying philosophy to be effective (Renger, Bartel & Foltysova 2013; Weiss 1995). An understanding of the within and between subsystem processes is critical in defining a meaningful efficiency goal. In most cases, the efficiency goal will emerge through discussions with SMEs while developing the process models, the topic of efficiency is inherent in the defining process.

Defining effectiveness goals is similar to program evaluation concepts of merit and worth, outcome evaluation, summative evaluation, and so forth (Greene 1994; Speer 1998; Cashin & Downey 1992). Abiding by the feasibility standard, SET recommends defining effectiveness goals by first examining any outcome data being collected within the system. Such data may be found within the system feedback mechanisms and/or other system data repositories. Systems often have built in *redundancies* therefore, it's important to examine each subsystem for data collection overlap (Denehy & Hsu 2003). Often areas of overlap represent a shared understanding of information deemed critical to establish success (Renger et al. in press a).

Engaging system and subsystem leadership as well as system actors is essential in defining efficiency and

effectiveness goals. Examining data gathered by existing feedback mechanisms can also provide the evaluator with valuable insight regarding possible efficiency and effectiveness goals (Vartanian 2010). If the system has built in mechanisms to monitor and assess data, it is assumed those data are potentially meaningful to the system leaders and operators. As efficiency and effectiveness standards are identified, they need to be carefully weighed for their potential to inform decision-making (LeCompte 2000). The concept of *representativeness* is important in ensuring the collected data will be useful in making system-wide decisions (LeCompte 2000).

Guiding Subprinciple 1i

Validate system definitions and goals

As system elements are defined, it is critical they be validated. Validation is necessary to ensure accuracy, buy-in and use (Sanders 1994). To the extent possible, leadership should also be engaged in validation. Validation methods include, but are not limited to, SME member checks (Renger & Bourdeau 2004); a review of source documentation (Renger 2011); a review of the literature (Foltysova 2013); inference from user-valued data elements, and so forth.

SET Guiding Principle 2

System efficiency is a necessary prerequisite for optimal system effectiveness

System efficiency depends on streamlined subsystem processes and coordination between subsystems (Comfort et al. 2004). To evaluate efficiency, it is necessary to assess: a) the health of feedback mechanisms; and, b) the extent to which within and between subsystem processes and system attributes are aligned to the efficiency goal.

Guiding Subprinciple 2a

Feedback mechanisms must provide timely, relevant, credible, frequent, and specific information to maximize efficiency

A systematic evaluation of feedback mechanisms must occur to ensure maximal system efficiency. First, it is necessary to establish whether the needed system feedback mechanisms exist. If they do exist, their functionality must be evaluated. To do this, each feedback mechanism should be evaluated using five criteria: timeliness, relevance, credence, frequency and specificity (Chen et al.



2014). Leadership and system actors must be engaged in defining each of these parameters or the feedback will not be used (Mohan 2014; Patton 2008).

Guiding Subprinciple 2b

Attributes must be aligned to maximize system efficiency

Misalignment of subsystem processes and attributes will result in waste and ultimately reduce system efficiency. To achieve alignment, a common goal is necessary. The importance of alignment to a common goal to increase efficiency and ultimately effectiveness rationale, mirrors the program evaluation concept of logic modeling, where alignment of activities to targeted conditions is a necessary prerequisite for program success (Renger & Titcomb 2002; Deyo, Julian & Jones 1995).

1. Alignment of system actor competence and capabilities

Societal systems and subsystems often provide training opportunities to ensure actors are competent and capable. Often these trainings occur in isolation, within a subsystem. Thus, it is possible that system-wide trainings will not be coordinated toward a common goal. One way to achieve coordination of effort is to work with training providers to ensure training elements are aligned to the efficiency goal. Program evaluation methods such as the logic model can be useful in ensuring training activities are meaningfully aligned to the efficiency goal (McLaughlin & Jordan 1999; Renger & Titcomb 2002).

2. Alignment of leadership

To maximize efficiency, system and subsystem leadership must have a shared and agreed upon understanding of a common goal. There is ample research describing the consequences when stakeholders do not have a shared understanding of the common goal (Coşkun, Akande, & Renger 2012; Innes 2004; Reed 2008). Examples of system inefficiencies include duplication of effort, unnecessary redundancies, waste, and so forth (Landau 1969). The efficiency goal can be useful in ensuring leadership is working strategically toward system streamlining and impact.

3. Alignment of information technology

In modern day systems, information technology (IT) is key to information flow and system efficiency. Information technology can be critical in developing efficient feedback mechanisms, streamlining competency and capability training for leadership communication, and general system functioning.

With technology comes many challenges, which can inadvertently undermine system and subsystem

efficiencies. For example, because of the silo in which subsystems often operate, they may be using different communication platforms, operating systems, databases, and so forth (Kasunic 2001). The evaluation must examine the compatibility of the IT infrastructure and seek to remedy interoperability challenges.

4. Alignment of culture

If there is cultural discord regarding the value of the goal, then the system will not realize its maximal efficiency. Alignment of culture requires stakeholders to be bought-in to the culture being espoused. The standards and phenomena desired need to be articulated by system leadership. One method of aligning culture is through the use of five business strategies designed to shift culture: 1) match strategy to culture; 2) focus on a few key aspects of culture to address; 3) use the strengths of existing culture; 4) integrate both formal (e.g. performance reviews) and informal (e.g. behavior modeling) interventions; and, 5) continually monitor and measure cultural change (Katzenbach, Steffen & Kronley 2012). Although a critical element to this process is leadership, an often-overlooked asset to ensuring alignment of culture is to utilize members and sub-groups who exemplify the desired culture change, to become champions of the desired culture.

Guiding Subprinciple 2c

Evaluate alternative pathways to improve efficiency

The system theory concept of the *reflex arc* is important in evaluating system efficiency. The stimulus response in most organisms requires the perception of the stimulus: the transfer of the signal to the brain through the sensory neurons and the return of the appropriate action through the motor neurons. In the reflex arc, the process is simplified and sped up—the sensory neuron goes to the spinal cord, is interpreted, and returned via a motor neuron (Phillips 1971); this is an example of how a system can adapt based on feedback, such adaptation results in an organism able to react more quickly to stimuli.

When evaluating efficiency, attention must be paid to the system's ability to detect and establish alternative pathways for improving efficiencies. Here, it is necessary to simultaneously consider the relationship between feedback mechanisms and process models. Credible and relevant feedback mechanisms will provide timely, specific information about corrective actions needed to improve the system processes. At a subsystem level, this could result in removing or combining process steps. At a system level, this could result in significant changes to the system structure. For example, based on credible, specific, and timely feedback, rural emergency response systems are changing system patient transport processes to bypass



smaller critical access hospitals and instead transfer directly to larger tertiary care facilities, thus saving time and cost, without sacrificing patient outcomes (i.e. effectiveness) (Liu, Bellamy, McCormick 2007).

SET Guiding Principle 3

Evaluate system effectiveness after evaluating system efficiency

Another important principle of system theory is that parts and subsystems are interdependent and work together *synergistically*; as such, systems are viewed *holistically*. Because of these system theory principles, SET recognizes the need to evaluate the effectiveness of the system as a whole—attempting to isolate and evaluate the effectiveness of any part or subsystem is nonsensical. Thus, the debate on what constitutes credible evidence is moot when conducting system evaluation (Donaldson & Lipsey 2006).

Engaging system and subsystem leadership as well as system actors is essential in defining possible effectiveness outcomes. Examining data gathered by system feedback mechanisms can also provide the evaluator with valuable insight regarding possible outcome measures. As potential effectiveness measures are identified, they need to be carefully weighed for their potential to inform decision-making (LeCompte 2000). Here again the concept of *representativeness* is important in ensuring the outcome data collected will be useful in making system-wide decisions.

Discussion

Our field is recognizing the potential of system thinking and system theory to advance the quality of evaluations. However, to date, the primary focus is on using these concepts to increase the quality of program evaluations; there is no guidance in the evaluation literature for evaluators to conduct evaluations of systems.

SET is offered as the first step forward to conducting system evaluations. The three SET guiding principles align with the basic tenets of system theory. The first principle is focused on defining the parts of the system. Principles two and three are built on the understanding that the defined elements must work together toward a common goal.

Limiting the number of SET guiding principles to three was purposeful and true to the concept of

parsimony. This is not to suggest the evaluation resulting from the application of these three principles will be simple. On the contrary, it is likely to be more complex and larger in scale than typical program evaluations. However, SET provides a necessary structure to work through the complexity and produce a meaningful evaluation with usable results, to untangle the wicked mess. The three guiding principles also provide a way to articulate a simply strategy for a complex and messy problem—a necessary element to secure the funding needed to conduct system evaluations from potential sponsors (Sanders 1994).

The three SET guiding principles were developed and tested in the health sector. Further testing of SET's utility in other contexts is needed. Nevertheless, SET represents an important first step in providing a blueprint for using system thinking and system theory to evaluate systems.

In developing SET, numerous evaluation and biology colleagues offered their critical review. One point of disagreement was the point at which goals should be defined. Some colleagues feel it is necessary to define the goals first. This would allow a more deliberate strategy in defining subsystems and the respective perspectives. For example, if the goal were to improve the efficiency and effectiveness of response to time-critical cardiac events, then the evaluation should include any subsystems and actors working toward that goal. Essentially, defining the goals helps establish the boundaries and perspectives. My concern with this approach is similar to that of retrofitting in logic modeling (Renger 2011). That is, only the perspectives of those who currently share an understanding of the goal will be included. There may be other critical subsystems, which should share the common goal, but do not, and therefore, are excluded from consideration.

From a methods standpoint, it appears that most of the program evaluation mixed-methods are useful when evaluating systems. Numerous SET concepts, such as aligning system attributes to an efficiency goal; engaging leadership in defining boundaries; and, recognizing different perspectives in defining processes, have parallel concepts in many program evaluation theories grounded in values, use, and methods (Alkin & Christie 2004). Table 1 depicts the relationship between SET principles, systems theory, and several evaluation methods. Undoubtedly, unique features of a system will dictate the development of new evaluation methods to meet the evaluation challenges posed by them.



TABLE 1. THE RELATIONSHIP BETWEEN SET PRINCIPLES, SYSTEMS THEORY, AND EVALUATION METHODS

SET principles	Relevant systems theory principle	Relevant evaluation standards and methods
<p>1. It is necessary to define the system, (including its parts, subsystems, relationships, attributes, inputs and common goal) before developing an evaluation plan.</p>	<p>Boundaries: parameters that limit the extent of the system—they may be physical, conceptual or mental (Ericson 2011).</p> <p>Perspectives: using different levels of system stakeholders to understand boundaries and processes (Williams & Hummelbrunner 2010).</p> <p>Subsystem: a major component of a system. It is made up of two or more interacting and interdependent components. Subsystems of a system interact in order to attain their own purpose(s) and the purpose(s) of the system in which they are embedded (Banathy 1992).</p> <p>Embeddedness: a state in which one system is nested in another system (Banathy 1992).</p> <p>Complexity: a systemic characteristic that stands for a large number of densely connected parts and multiple levels of embeddedness and entanglement. Not to be confused with complicatedness, which denotes a situation or event that is not easy to understand, regardless of its degree of complexity (Laszlo 2009).</p> <p>Environment: the context within which a system exists. It is composed of all things that are external to the system, and it includes everything that may affect the system, and may be affected by it at any given time (Ericson, 2011).</p> <p>Feedback: is a process by which information concerning the adequacy of the system, its operations and its output, are introduced into the system (Banathy 1992).</p> <p>Process models: a description of process. Most often a process model is used for a descriptive purpose, that is, to describe ‘how things must/should/could/be done’ in contrast to the process itself, which is what really happens (Rolland 1993).</p> <p>Attributes: is a quantitative or qualitative characteristic of a system or system element; for example, safety, reliability and quality are systems attributes that result as an emergent property of the final system design (Ericson 2011).</p> <p>Inputs: what is put in, taken in, or operated on by any process or system (Oxford Dictionaries online 2015).</p>	<ul style="list-style-type: none"> ■ Utilization standard engaging system leaders to endorse evaluation ■ Qualitative Interview Methods ■ Storyboarding
<p>2. System efficiency is a necessary prerequisite for optimal system effectiveness.</p>	<p>Cascading failure: is the failure scenario in a system of interconnected parts whereby the failure of an item directly causes another item to fail (Ericson 2011).</p> <p>Feedback mechanisms: provide for the analysis and interpretation of information generated about the adequacy of the output and the operations of the system. The interpretation is used for designing and introducing adjustments (into the system) in order to bring about adequate output and improved systems operations (Banathy 1992).</p>	<ul style="list-style-type: none"> ■ Feasibility standard: Using system attributes to also serve an evaluation function ■ Continuous quality improvement ■ Process flow mapping ■ Emergency preparedness exercises ■ Logic modeling
<p>3. Evaluate system effectiveness after evaluating system efficiency.</p>	<p>Holism: attention is focused on the emergent properties of the whole rather than on the reductionist behavior of the isolated parts. The approach typically involves and generates empathetic, experiential, and intuitive understanding, not merely analytic understanding (Ericson 2011).</p> <p>Synergy: is the process by which a system generates emergent properties resulting in the condition in which a system may be considered more than the sum of its parts, and equal to the sum of its parts plus their relationships. This resulting condition can be said to be one of synergy (Ericson 2011).</p>	<p>Credible evidence debate: Cause and effect analysis NOT relevant</p>



While programs vary in scale, it is reasonable to posit most systems and the accompanying evaluation will be larger in scale than what is typically encountered in a program evaluation (McDavid, Huse & Hawthorn 2013). As the number of subsystems and processes increases, so does the scale and complexity of the evaluation. The reality is evaluators will likely have insufficient budgets to evaluate all aspects of system efficiency and effectiveness. This reality alone may explain why the evaluation literature is devoid of articles devoted to conducting systems evaluation, and to my knowledge, why no one has ever been provided the funding to actually evaluate an entire system. The flexibility to explore evaluation options was only made possible through the generosity of the Leona M. and Harry B. Helmsley Charitable Trust.

One strategy to reduce evaluation cost is to recognize the duality of system attributes; that is, the same attributes needed for system functioning should, if possible, be leveraged to complete the evaluation. For example, leadership is a core system attribute necessary for human systems to function efficiently and effectively. However, leadership is also a necessary attribute for the success of the system evaluation (Patton 2008). Similarly, IT is key to the efficient and effective functioning of many systems. The evaluator with an eye toward feasibility will examine ways to exploit existing IT infrastructure to also meet evaluation needs.

SET is the first attempt to provide a blueprint for guiding evaluators through the complexity of evaluating systems. As SET guiding principles are tested in evaluating other systems, other evaluators will hopefully assess the robustness of the principles and make suggestions to improve application.

Acknowledgements

The author would like to acknowledge the contributions of Gabriele Bartel, May Bjerke, Dr. Jirina Foltysova, Trista Kontz-Bartels, Halley Lee, Marty Link, Makenzie McPherson, Tom Nehring, Jessica Renger, Dr. Carlos Rodriguez, and Eric Souvannasac— their insights and feedback have been invaluable.

Endnote

- 1 As noted earlier, herein lies a possible source of confusion. In SET the evaluation focuses on establishing the quality of the existing system feedback mechanism. This is not to be confused with the feedback provided as a result of the evaluation.

References

Abdulmalek, F & Rajgopal, J 2007, 'Analyzing the benefits of lean manufacturing and value stream mapping via simulation: a process sector case study', *International Journal of Production Economics*, vol. 107, no. 1, pp. 223–236, DOI:10.1016/j.ijpe.2006.09.009.

- Adair, J 2009, *Not bosses but leaders how to lead the way to success*, 3rd edn, Kogan Page, London.
- Adams, KM, Hester, PT, Bradley, JM, Meyers, TJ & Keating, CB 2014, 'Systems theory as the foundation for understanding systems', *Systems Engineering*, vol. 17, no. 1, pp. 112–123, DOI: 10.1002/sys.21255.
- Alkin, M & Christie, C 2004, 'An evaluation theory tree', in *Evaluation roots: tracing theorists' views and influences*, 1st edn, pp. 12–65, Sage Publications, Thousand Oaks, California.
- Andrews, HA 1985, *Evaluating for excellence: addressing the need for responsible and effective faculty evaluation*, New Forums Press, Inc., Stillwater, Oklahoma.
- Arvey, RD, Campbell, JP, Dunnette, MD & Hellervik, LV 1973, 'The development and evaluation of behaviorally based rating scales', *Journal of Applied Psychology*, vol. 57, no. 1, pp. 15, DOI: 10.1037/h0034185.
- Baker, AH & Manoj, BS 2007, 'Communication challenges in emergency response', *Communications of the ACM*, vol. 50, no. 3, pp. 51–53, DOI: 10.145/1226736.1226765.
- Bakos, JY & Treacy, ME 1986, 'Information technology and corporate strategy: a research perspective', *MIS Quarterly*, vol. 10, no. 2, pp. 107–119, DOI: 10.2307/249029.
- Banathy, BH 1992, *A systems view of education: concepts and principles for effective practice*, Educational Technology, Englewood Cliffs, New Jersey.
- Barrett, G & Odum E 2004, *Fundamentals of ecology*, 5th edn, Brooks Cole, Belmont, California.
- Becker, K, Renger, R & McPherson, M 2015, 'Indicators of buy-in to gauge evaluation success', *Australasian Journal of Evaluation*, vol. 15, no. 2, pp. 12–21.
- Borich, GD & Jemelka, RP 1982, *Programs and systems: an evaluation perspective*, Academic Press, San Diego, CA.
- Boyle, R & Lemaire, D 1999, *Building effective evaluation capacity: lessons from practice*, vol. 8, Transaction Publishers, New Brunswick, New Jersey.
- Brandon, RN 2014, *Adaptation and environment*, Princeton University Press, Princeton, New Jersey.
- Buckley, WF 1998, *Society—a complex adaptive system: essays in social theory*, Gordon and Breach, Australia.
- Buckley, WF 1967, *Sociology and modern systems theory*, Prentice-Hall, Englewood Cliffs, New Jersey.
- Burns, TR 2007, 'System theories', in G Ritzer (ed), *The Wiley-Blackwell Encyclopedia of Sociology*, Malden, MA and Blackwell Publishing, Oxford.
- Buzna, L, Helbing, D & Peters, K 2008, 'Modelling of cascading effects and efficient response to disaster spreading in complex networks', *International Journal of Critical Infrastructures*, vol. 4, nos. 1–2, pp. 46–62, DOI: 10.1504/UCIS.2008.016091.
- Cashin, WE & Downey, RG 1992, 'Using global student rating items for summative evaluation', *Journal of Educational Psychology*, vol. 84, no. 4, pp. 563.
- Checkland, P & Poulter, J 2010, 'Soft systems methodology', in *Systems approaches to managing change: a practical guide*, 1st edn, pp. 191–242, Springer, London.
- Checkland, P 1999, *Systems thinking. Rethinking management information systems*, Oxford University Press, New York.
- Chen, H, Hailey, D, Wang, N & Yu, P 2014, 'A review of data quality assessment methods for public health information



- systems', *International Journal of Environmental Research and Public Health*, vol. 11, no. 5, pp. 5170–5207.
- Christie, CA & Alkin, MC 2008, 'Evaluation theory tree reexamined', *Studies in Educational Evaluation*, vol. 34, no. 3, pp. 131–135.
- Cohen, E & Tichy, NM 2003, 'Why are leaders important?', in *Business Leadership: A Jossey-Bass reader*, pp. 4–28, Jossey-Bass, San Francisco.
- Cohen, AY & Kibel, BM 1993, *The basics of open systems evaluation: a resource paper*, The Pacific Institute for Research and Evaluation, Chapel Hill, North Carolina.
- Comfort, LK, Dunn, M, Johnson, D, Skertich, R & Zagorecki, A 2004, 'Coordination in complex systems: increasing efficiency in disaster mitigation and response', *International Journal of Emergency Management*, vol. 2, nos. 1–2, pp. 62–80.
- Coşkun, R, Akande, A & Renger, R 2012, 'Using root cause analysis for evaluating program improvement', *Evaluation Journal of Australasia*, vol. 12, no. 2, pp. 4–14.
- Curphy, GJ, Hogan, J & Hogan, R 1994, 'What we know about leadership: effectiveness and personality', *American Psychologist*, vol. 49, no. 6, pp. 493.
- Denehy, TE & Hsu, WW 2003, *Duplicate management for reference data*, research report RJ10305, IBM Research.
- Department of Homeland Security 2013, *Homeland Security Exercise and Evaluation Program*, Federal Emergency Management Agency.
- Deyo, D, Julian, DA & Jones, A 1995, 'Open systems evaluation and the logic model: program planning and evaluation tools', *Evaluation and Program Planning*, vol. 18, no. 4, pp. 333–341.
- Donaldson, SI & Lipsey, MW 2006, 'Roles for theory in contemporary evaluation practice: developing practical knowledge', in I Shaw, JC Greene & MM Mark (eds), *The handbook of evaluation: policies, programs, and practices*, pp. 56–75, Sage Publications, London.
- Donnelly, J 1984, 'Cultural relativism and universal human rights', *Human Rights Quarterly*, vol. 6, no. 4, pp. 400–419.
- Eisenberg, MS 2013, *Resuscitate: how your community can improve survival from sudden cardiac arrest*, University of Washington Press, Washington.
- Ericson, CA 2011, *Concise encyclopedia of system safety: definition of terms and concepts*, John Wiley & Sons, Hoboken, New Jersey.
- Fainter, J 1991, 'Quality assurance not equal to quality improvement', *Journal of Quality Assurance*, vol. 13, no. 1, pp. 8–9.
- Fetterman, DM 1994, 'Empowerment evaluation', *Evaluation Practice*, vol. 15, no. 1, pp. 1–15.
- Flynn, BB, Schroeder, RG & Sakakibara, S 1994, 'A framework for quality management research and an associated measurement instrument', *Journal of Operations Management*, vol. 11, no. 4, pp. 339–366.
- Fullan, M 2004, *Leadership & sustainability: system thinkers in action*, Corwin Press, Thousand Oaks, California.
- Foltysova, J 2013, *Validation of reconstructed program theory*, University of Arizona, Tucson, Arizona.
- Gershon, R, Stone, P, Bakken, S & Larson, E 2004, 'Measurement of Organizational Culture and Climate in Healthcare', *The Journal of Nursing Administration*, vol. 34, no. 1, pp. 33–40.
- Gibson, JJ 1960, 'The concept of the stimulus in psychology', *American Psychologist*, vol. 15, no. 11, pp. 694.
- Greene, JC 1994, 'Qualitative program evaluation: practice and promise', in NK Denzin & YS Lincoln (eds), *Handbook of qualitative research*, Sage, Thousand Oaks, California, pp. 530–544.
- Heifetz, RA & Laurie, DL 1997, 'The work of leadership', *Harvard Business Review*, vol. 75, no. 1, pp. 124–134.
- Holweg, M 2007, 'The genealogy of lean production', *Journal of Operations Management*, vol. 25, no. 2, pp. 420–437, DOI:10.1016/j.jom.2006.04.001.
- Innes, JE 2004, 'Consensus building: Clarifications for the critics', *Planning Theory*, vol. 3, no. 1, pp. 5–20.
- Jeliazkova, M & Westerheijden, DF 2002, 'Systemic adaptation to a changing environment: towards a next generation of quality assurance models', *Higher Education*, vol. 44, nos. 3–4, pp. 433–448.
- Juergens, HF 1977, 'Attributes of information system development', *MIS Quarterly*, vol. 1, pp. 31–41.
- Kasunic, M 2001, *Measuring systems interoperability: challenges and opportunities*, Carnegie-Mellon University Pittsburgh Software Engineering Institute, Pittsburgh, Pennsylvania.
- Katzenbach, JR, Steffen, I & Kronley, C 2012, 'Cultural change that sticks', *Harvard Business Review*, vol. 90, no. 7, pp. 8.
- Kazdin, AE 1979, 'Unobtrusive measures in behavioral assessment', *Journal of Applied Behavior Analysis*, vol. 12, no. 4, pp. 713–724.
- Kotter, JP 2001, 'What leaders really do', *Harvard Business Review*, vol. 71, no. 11, pp. 3–11.
- Landau, M 1969, 'Redundancy, rationality, and the problem of duplication and overlap', *Public Administration Review*, vol. 1, no. 4, pp. 346–358.
- Lanzendorf, M, Rubens, R, Vahlhaus, M & Zintl, M 2011, 'Final panel discussion: conclusions', paper presented at the Deutsche Gesellschaft für Internationale Zusammenarbeit Systemic Approaches in Evaluation Conference, Eschborn, Germany, January.
- Laszlo, A 2009, 'The nature of evolution', *World Futures*, vol. 65, no. 3, pp. 204–221.
- LeCompte, MD 2000, 'Analyzing qualitative data', *Theory into Practice*, vol. 39, no. 3, pp. 146–154.
- Liu, JJ, Bellamy, GR & McCormick, M 2007, 'Patient bypass behavior and critical access hospitals: implications for patient retention', *Journal of Rural Health*, vol. 23, no. 1, pp. 17–24.
- Mark, MM, Henry, GT & Julnes, G 2000, *Evaluation: an integrated framework for understanding, guiding, and improving policies and programs*, Jossey-Bass, San Francisco, California.
- McDavid, JC, Huse, I & Hawthorn, LRL 2013, 'Key concepts and issues in program evaluation and performance measurement', in *Program evaluation and performance measurement: an introduction to practice*, 2nd ed., Sage, Thousand Oaks, California.
- McLaughlin, JA & Jordan, GB 1999, 'Logic models: a tool for telling your programs performance story', *Evaluation and Program Planning*, vol. 22, no. 1, pp. 65–72.
- Meadows, DH & Wright, D 2008, *Thinking in systems: a primer*, Chelsea Green Pub, White River Junction, Vermont.
- Miller-Williams, SL & Kritsonis, WA 2009, 'A systems approach to comprehensive school reform: using the realms of



- meaning and the Baldrige model as a systems framework', *National Forum of Applied Educational Research Journal*, vol. 23, nos. 1–2, pp. 1–8.
- Mohan, R 2014, 'Evaluator advocacy it is all in a day's work', *American Journal of Evaluation*, vol. 35, no. 3, pp. 397–403.
- Morell, JA 2005, 'Why are there unintended consequences of program action, and what are the implications for doing evaluation?', *American Journal of Evaluation*, vol. 26, no. 4, pp. 444–463.
- Odum, E 1971, *Fundamentals of ecology*, 3rd edn, Saunders, Philadelphia, Pennsylvania.
- Øvretveit, J & Gustafson, D 2002, 'Evaluation of quality improvement programmes', *Quality and Safety in Health Care*, vol. 11, no. 3, pp. 270–275.
- Oxford Dictionaries online, 2015, Input definition, in Oxford Dictionaries online, Oxford University Press, viewed 4 July 2015, http://www.oxforddictionaries.com/us/definition/american_english/input.
- Parsons, T 1961, 'An outline of the social system', *Classical Sociological Theory*, vol. 2, pp. 421–440.
- Patton, MQ 2008, *Utilization-focused evaluation*, Sage Publications, Thousand Oaks, California.
- Phillips, DC 1971, 'James, Dewey, and the reflex arc', *Journal of the History of Ideas*, vol. 32, no. 4, pp. 555–568.
- Radnor, Z & Walley, P, Stephens, A & Bucci, G 2006, *Evaluation of the Lean approach to business management and its use in the public sector*, Scottish Executive Social Research, Edinburgh, <http://www.gov.scot/Resource/Doc/129627/0030899.pdf>.
- Reed, GE 2006, 'Leadership and systems thinking', *Defense AT&L*, vol. 35, no. 3, pp. 10–13.
- Reed, MS 2008, 'Stakeholder participation for environmental management: a literature review', *Biological Conservation*, vol. 141, no. 10, pp. 2417–2431.
- Renger, R, Foltysova, J, Becker, K & Souvannasacd, E (in press a), 'The power of the context map: designing realistic outcome evaluation strategies and other unanticipated benefits', *Evaluation and Program Planning*, vol. 52, pp. 118–125.
- Renger, R, Becker, K, Rogan, M & Kontz-Bartels, T (in press b), 'Process flow mapping for systems improvement: lessons learned', *Canadian Journal of Program Evaluation*.
- Renger, R, Bartel, G & Foltysova, J 2013, 'The reciprocal relationship between implementation theory and program theory in assisting program design and decision-making', *Canadian Journal of Program Evaluation*, vol. 28, no. 1, pp. 27–41.
- Renger, R, Wood, S, Williamson, S & Krapp, S 2012, 'Systemic evaluation, impact evaluation and logic models', *Evaluation Journal of Australasia*, vol. 11, no. 2, pp. 24–30.
- Renger, R 2011, 'Constructing and verifying program theory using source documentation', *The Canadian Journal of Program Evaluation*, vol. 25, no. 1, pp. 51–67.
- Renger, R & Bourdeau, B 2004, 'Strategies for values inquiry: an exploratory case study', *American Journal of Evaluation*, vol. 25, no. 1, pp. 39–49.
- Renger, R & Titcomb, A 2002, 'A three-step approach to teaching logic models', *American Journal of Evaluation*, vol. 23, no. 4, pp. 493–503.
- Rogers, PJ 2011, 'Program theory and logic models for systemic evaluation', paper presented at the Deutsche Gesellschaft für Internationale Zusammenarbeit Systemic Approaches in Evaluation Conference, January, Eschborn, Germany.
- Rogers, PJ 2008, 'Using programme theory to evaluate complicated and complex aspects of interventions', *Evaluation*, vol. 14, no. 1, pp. 29–48.
- Rolland, C 1993, 'Modeling the requirements engineering process', paper presented at the 3rd European-Japanese Seminar on Information Modelling and Knowledge Bases, Budapest, Hungary.
- Sanders, JR 1994, *The program evaluation standards: how to assess evaluations of educational programs*, 2nd edn, Sage Publications, Thousand Oaks, California.
- Schein, EH 1990, 'Organizational culture', *American Psychologist*, vol. 45, no. 2, pp. 109–119, DOI:10.1037/0003-066X.45.2.109.
- Schwab, DP, Heneman III, HG & DeCotiis, TA 1975, 'Behaviorally anchored rating scales: a review of the literature', *Personnel Psychology*, vol. 28, pp. 549–562.
- Scott, T, Mannion, R, Davies, H & Marshall, M 2003, 'The quantitative measurement of organizational culture in health care: a review of the available instruments', *Health Services Research*, vol. 38, no. 3, pp. 923–945, DOI: 10.1111/1475-6773.00154.
- Speer, DC 1998, *Mental health outcome evaluation*, Academic Press, San Diego, California.
- Stame, N 2004, 'Theory-based evaluation and types of complexity', *Evaluation*, vol. 10, no. 1, pp. 58–76.
- Vartanian, TP 2010, *Secondary data analysis*, Oxford University Press, New York.
- Von Bertalanffy, L 1968, *General system theory: foundations, development, applications*, George Braziller, New York.
- Weiss, CH 1995, 'Nothing as practical as good theory: exploring theory-based evaluation for comprehensive community initiatives for children and families', in JP Connell, AC Kubisch, LB Schorr, CH Weiss (eds), *New approaches to evaluating community initiatives. Concepts, methods, and contexts*, Aspen Institute, Washington DC.
- Williams, JC 2015, 'A systems thinking approach to analysis of the Patient Protection and Affordable Care Act', *Journal of Public Health Management and Practice*, vol. 21, no. 1, pp. 6–11, DOI: 10.1097/PHH.0000000000000150.
- Williams, B & Hummelbrunner, R 2010, *Systems concepts in action: a practitioner's toolkit*, Stanford University Press, Palo Alto, California.



**BRIDGET DOHERTY | RICHARD ECCLESTON | EMILY HANSEN | KRISTIN NATALIER |
BRENDAN CHURCHILL**

Building evaluation capacity in micro community organisations—more burden than benefit?

Building evaluation capacity in community organisations is promoted as a key means of creating an evaluation culture that will reorient organisations towards identifying and measuring outcomes, demonstrating achievements and enhancing accountability.

Building evaluation capacity presents a number of challenges for micro-organisations, including limited resources for evaluation, a low existing evaluation skills base, and, an organisational culture and practice that does not align well with evaluation tools and techniques typically used to demonstrate program outcomes. The implementation of the Thriving Communities Healthy Families pilot project in Tasmania, a place-based health promotion intervention delivered through Neighbourhood Houses (NH), included an evaluation capacity building (ECB) component. The pilot project is used in this article as a case study to illustrate the tensions that arose for community development workers in the NH as they implemented a new program, and contended with the demands of developing their evaluation capacity. We found that while community development workers value evaluation and the importance of demonstrating program outcomes, the model of building evaluation capacity of the workers themselves was burdensome. Further, some workers perceived that evaluation activities impacted negatively on their primary role of supporting vulnerable community members.

We conclude that, in very small community sector organisations, ECB requires an ongoing mentoring/relationship-building approach, rather than a traditional training approach. This has implications for how to fund ECB and raises questions on where best to focus ECB in these micro-organisations.

Introduction

The Australian community, or not-for-profit (NFP), sector is a diverse and complex system in which contexts, client needs and circumstances are continuously changing (ARACY 2009). The NFP sector, both in Australia and abroad, is playing an increasingly central role in the delivery of human services. There are many explanations for this phenomenon, including the cost effectiveness and expertise within the NFP sector and, in the case of micro-organisations (which are the focus of this article), their strong connections to the communities they serve (McCoy, Rose & Connolly 2013; Productivity Commission 2010; Ernst & Young 2014; DHHS 2014). As NFPs become increasingly important agents of service delivery, governments and funding bodies are demanding they demonstrate the effectiveness of their programs and



Bridget Doherty is a Research Manager at the Institute for the Study of Social Change at the University of Tasmania.

Email: bridget.doherty@utas.edu.au

Richard Eccleston is Director at the Institute for the Study of Social Change at the University of Tasmania.

Email: richard.eccleston@utas.edu.au

Emily Hansen is a Lecturer in Sociology at the School of Social Sciences, University of Tasmania.

Email: emily.hansen@utas.edu.au

Kristin Natalier is an Associate Professor in Sociology at the School of Social and Policy Studies, Flinders University.

Email: kris.natalier@flinders.edu.au

Brendan Churchill is an Associate Lecturer in Sociology at the School of Social Sciences, University of Tasmania.

Email: brendan.churchill@utas.edu.au

their broader contribution to society. This emerging ‘culture of audit’ and associated demands for evidence of policy impact, places new demands on NFPs, including the increasingly common expectation that NFPs are able to develop their own evaluation capacity. Given this context, we are especially interested in how NFP micro-organisations respond to this challenge.

Preskill and Boyle (2008) argue that the growing need to embark upon evaluation capacity building (ECB) in the community sector is being driven by both external and internal pressures. There is an external demand for accountability and evidence of outcomes to support decision-making, as well as pressures from within organisations to build new skills and knowledge, improve programs and services, and maximise new funding opportunities. We recognise the clear need to demonstrate the efficacy of social programs, but we argue that building evaluation capacity presents a number of challenges, particularly when micro-organisations have limited resources for evaluation and a low existing evaluation skills base. Further, we argue that the prevailing community development culture and model of practice within many NFP micro-organisations does not sit easily with the evaluation tools and techniques typically used to demonstrate program outcomes.

In this article, we highlight the tensions and challenges of building evaluation capacity in NFP micro-organisations through a case study—the Thriving Communities Healthy Families (TCHF) project; an innovative place-based health promotion initiative in Tasmania. We provide a brief overview of the broader policy context (within which the TCHF project sits) including: the evidence-based policy movement; the growth of performance auditing (with its emphasis on outcomes rather than outputs); and, the place-based policy approach. We then use the case study to illustrate the challenges of ECB in NFP micro-organisations and conclude by outlining strategies that may enable actors at the front line of human services delivery to better meet the evaluation challenge they are facing.

The origins of the evidence-based policy movement

The push for evidence-based social policy is generally agreed to have emerged out of the Blair Government (UK) of the late 1990s. Solesbury (2001) points to other influences including the gearing of research funding to governments’ social policy priorities; making academic research more user-friendly; and an increasing interest from practitioners themselves in demonstrating the efficacy of their own practices. These new influences signalled a shift towards a more pragmatic ‘what works is what matters’ approach (Solesbury 2001; Marston & Watts 2003; Banks 2009).

As Banks (2009) notes, evidence and analysis would ideally inform each stage in the policy process—from issue identification through to policy options; this process would mitigate the influence of sectional and political interests and allow for better community consideration of potential impacts. However, the policy process is more likely to be driven by political imperatives and short timeframes that do not fully allow for such considerations. Evidence is but one ‘input’ into the process and what constitutes evidence is contested (Bacchi 2009). There is a risk that only privileged forms of evidence (that of experts or derived from scientific methodologies) will be accepted, and other forms of knowledge (such as lay knowledge) will be excluded (Marston & Watts 2003). This is an important consideration for place-based interventions like the TCHF, that look to the community as an important source of policy-relevant knowledge and use a range of methods (often time-consuming) to access that knowledge.

Evidence-based policy is a firm feature of Australian public policy making discourse and practice. However, Banks (2009) suggests that using an evidence-based approach poses a number of challenges for



government—including a dearth of robust data to inform program evaluation (particularly in social policy arenas) and a need to build research and evaluation capability and expertise. Banks suggests a need for more open and transparent policy formulation, as well as scrutiny of the methodologies used to produce evidence. Further, evaluation should be understood as a necessity rather than a luxury and should be specifically resourced from within program funding. Finally, and perhaps most importantly, there is a need to build in the necessary time to accumulate and test evidence.

The evidence-based policy approach is not limited to the evidence that informs policy formulation, that is, it also includes evidence for assessing the efficacy of program implementation and demonstrating outcomes (McCoy, Rose & Connolly 2013; Australian Charities and Not-for-profits Commission 2014). Such a need for evidence drives demand for more sophisticated and meaningful evaluation. Increasingly, reporting outcomes are a condition of the funding agreements between governments and the community sector organisations they contract to deliver human services programs (DHHS 2014). This trend for enhanced accountability places additional burdens on organisations in terms of data collection and management, reporting and compliance, and evaluation capacity.

The accountability challenge

Community sector organisations are generally distinguished from government and private sector entities by their social or community purposes, self-governance models and use of volunteers (ARACY 2009). These organisations operate under different constraints to the government or the for-profit sector and are mission oriented, focussing on providing benefits for the community and improving people's lives (Edwards 2012). This sector is largely concerned with the delivery of a wide range of services (including health and human services) that the private or government sectors do not, or cannot provide—for a range of financial, political and ideological reasons.

Organisations that are established for a community purpose operate differently in terms of how they are organised, how they engage with stakeholders, and how they make decisions and deliver services. These organisations value participatory and inclusive processes that can be time-consuming and thus costly (Productivity Commission 2010). The organisations' relationships with the community are trust-based and go beyond the delivery of services; their inclusive approach empowers families, creates social inclusion, generates community connections and strengthens civil society (Productivity Commission 2010), the very features that the TCHF project wants to harness.

Most community-based organisations receive a significant portion of their funding from governments

who require evidence of value for expenditure of public money. Likewise, donors and volunteers have an interest in seeing their resources used effectively. Thus, there is increasing pressure from governments, funding bodies and the public to 'demonstrate, and document, the differences [NFP] services are making for individuals, families and communities' (McCoy, Rose & Connolly 2013 p. 15). Where once NFPs enjoyed a special status by virtue of their self-evident 'good' work, this status is no longer unquestioned and increasingly, there is a call for them to demonstrate the effectiveness of their work (Hoole & Patterson 2008).

The push for greater accountability imposes 'costs' on NFPs, including evaluation and planning costs, which can be a source of tension between organisations and their funders. While most community sector organisations acknowledge the need for accountability, many feel that reporting requirements, micromanagement, restrictions on the range of services provided, and greater compliance burdens are disproportionate to organisational size and scale (Productivity Commission 2010). A 2014 Australian Charities and Not-for-profits Commission report noted that Commonwealth government-funded charitable organisations believed that regulatory and reporting costs had increased over the previous three years—with the greatest burden arising from obligations related to their funding agreements, rather than any legislative requirements. Further, organisations perceived the frequency of reporting and level of detail required as excessive, and more concerned with risk management than performance outcomes (Australian Charities and Not-for-profits Commission 2014).

From outputs to outcomes

Nationally and internationally, there is an emphasis on measuring client outcomes in the human services (at both individual and community levels), as well as a concomitant shift away from a focus on organisational activities and outputs. While outputs measure the efficacy of program delivery, outcomes measure program impacts on clients and the community (ARACY 2009). The shift to outcomes funding is often couched in terms of making organisations more client-focussed, integrated, and efficient, as opposed to being constrained by traditional bureaucratic and siloed funding arrangements (DHHS 2014). The focus on outcomes also aims to strengthen co-production and collaboration between funding bodies (government) and the community sector, while also providing a guide to the investment of scarce resources to achieve desired results. Advocates of outcomes-based accountability argue that it encourages organisations to be innovative, and positions them to better attract additional funding through demonstration of their results (DHHS 2014). Edwards (2012) describes outcomes as the NFP sector's equivalent of the for-profit industry's return on investment; the outcomes allow organisations to



report the changes they are making regarding a client or community's condition to their funders and stakeholders.

The Tasmanian Government (like other jurisdictions across Australia and beyond) is currently implementing an Outcomes Purchasing Framework (OPF) to manage its service agreements with approximately 250 community service organisations that provide health and human services throughout the state (DHHS 2014). This framework will replace existing compliance-focussed instruments and represents a significant shift in the relationship between the government and its funded organisations—from one of purchaser/provider to that of partners in defining and delivering outcomes (TasCOSS, 2014). Central to the OPF is a mutually negotiated 'Commissioning for Outcomes' statement, which defines performance indicators, targets and a monitoring process. The Tasmanian Council of Social Service (TasCOSS) supports the move toward measuring outcomes, with the caveat that this must be 'appropriately resourced and properly done' (TasCOSS 2014, p. 16).

Edwards (2012) suggests measuring and reporting outcomes is just the first step in a continuous process, which also includes regular monitoring and analyses of outcome data to improve and manage services. However, this process presents a number of challenges, not least the existing capacity of community sector organisations to collect and manage information, analyse data, and report and measure outcomes. One of the solutions posited to help such organisations meet these rising demands is to build organisational evaluation capacity (Suarez-Balcazar & Taylor-Ritzler 2013), however, this poses its own challenges for resource-limited, time poor and community-oriented organisations.

Evaluation capacity building

There are a number of definitions of ECB. Stockdill, Baizerman and Compton define ECB as:

A context-dependent, intentional action system of guided processes and practices for bringing about and sustaining a state of affairs in which quality program evaluation and its appropriate uses are ordinary and ongoing practices within and/or between one or more organizations/ programs/sites (2002, p. 8).

ECB includes training and skills building, evaluation practice, reflective practice, self-assessment, and organisational learning (Suarez-Balcazar & Taylor-Ritzler 2013; Beere 2005; Bourgeois & Cousins 2013). The ultimate goal is sustainable evaluation practice where 'members continuously ask questions that matter, collect, analyse, and interpret data, and use evaluation findings for decision-making and action' (Preskill & Boyle 2008, p. 444).

Bourgeois and Cousins (2013, p. 304) divide ECB into two broad categories: the 'capacity to *do*' and 'capacity to

use' evaluation. The former includes the human resources of the evaluation unit, namely leadership, budget and data collection systems, and evaluation planning and activities. The latter includes evaluation literacy, integration with organisational decision-making, and learning benefits. Bourgeois and Cousins suggest that community sector organisations tend to have higher capacity to 'use' rather than 'do' evaluation, due in part to their smaller scope of practice, which makes a stand-alone evaluation resource/function unlikely. In these organisations, managers are more likely to take on the evaluation role and are therefore, more likely to integrate evaluation into organisational decision-making.

There are a number of approaches (formal and informal) to building evaluation capacity in an organisation including introduction to evaluation tools and training, workshops, technical assistance, mentoring, and action research and consultation (Garcia-Iriarte et al. 2011; Preskill & Boyle 2008). Huffman, Thomas and Lawrenz (2008) criticise these approaches as typically focussing on building individual rather than organisational capacity. As an alternative to building the skills of individuals, Huffman et al propose a 'collaborative immersion' approach, whereby collaborative evaluation experiences in complex real world situations become the mechanism for building evaluation capacity, in both individuals and organisations. Irrespective of the approach taken, ECB strategies should take into account the organisation's characteristics, organisational resources, existing evaluation experience, as well as the desired learning objectives. It is also important to assess an organisation's readiness to engage in an ECB process (Garcia-Iriarte et al. 2011); this includes partners sharing and clarifying motivations, assumptions, and expectations. In a similar vein, Preskill and Boyle (2008) say it is important to explore organisational assumptions prior to embarking on ECB. These might include the assumptions that evaluation is self-evidently a 'good' thing—improving decision-making, adding value to the organisation, and encouraging staff members to want to learn and use evaluation skills in their practice.

While ECB can help organisations to meet accountability demands, the reality of small organisations' lack of resources, time and funding can impact negatively on ECB efforts. Garcia-Iriarte et al. (2001) posit a 'catalyst' model to address this, building ECB initially with one staff member, who in turn transfers learning to another, thus building up organisational capacity. It is important that the catalyst is an organisational leader who can facilitate learning in other staff members and mainstream the use of evaluation into organisational practice. To build sustainability, the organisation needs to internalise evaluation processes, systems, policies, and procedures that are self-renewing and evolving—taking a 'holistic, proactive, and conscious approach to evaluation' (Preskill & Boyle 2008, p. 454).



Preskill and Boyle (2008) distinguish between evaluation and ECB, describing them as different processes. However, as Beere (2005) notes, evaluators are often best positioned to do the work of ECB in organisations. Good program evaluation serves to model good evaluation practices by demonstrating how evaluation contributes to program and organisational development; this can build capacity and help overcome any negativity associated with evaluation. ECB is a continuous project, an ongoing process—just as evaluation itself is a process rather than an event (Beere 2005). The aim is to create an evaluation culture in organisations so that evaluation becomes a core business activity.

The Thriving Communities Healthy Families pilot project

The implementation of ECB in the Thriving Communities Healthy Families (TCHF) project illustrates the issues and challenges outlined above. The TCHF is a two-year pilot program to link disenfranchised, at-risk families with their local Neighbourhood (community) House (NH), located in five socio-economically disadvantaged areas throughout Tasmania. The project aims to improve families' access to services and programs, promote social inclusion, empower participants, and enhance health and wellbeing. NHs are micro-organisations; they have a volunteer board and generally have less than five paid employees, including a house coordinator and community development worker(s), many of whom work part time. Neighbourhood Houses Tasmania (NHT) is the peak body for the 33 NHs funded by the Tasmanian Government (Neighbourhood House Tasmania 2015). The TCHF project appointed a part-time Healthy Families worker in each house to develop and implement engagement strategies to identify and connect young families to the NH and wider community, build capacity, and, through a partnership model, connect the families to local health and wellbeing services.

The place-based approach, which draws on a 'social determinants of health' model, underpins the TCHF. This approach recognises that social, economic and environmental factors shape health outcomes and health behaviours (Dahlgren & Whitehead 1991), and 'place' mediates these outcomes. Place has a significant impact on how health inequities are manifested and produced. Australians living in socio-economically disadvantaged areas experience poorer outcomes across a range of health status indicators, including: mortality, morbidity, life expectancy, health risk behaviour and self-assessed health (Australian Bureau of Statistics 2012).

Place-based initiatives (PBI) are one approach used by governments to address health inequity. PBIs 'seek to improve the social, cultural, economic and/or physical environment within a defined boundary' (Boyd 2008). They target communities rather than individuals or

high-risk groups, and recognise that as localities differ, 'each will raise unique solutions' (The Royal Children's Hospital Centre for Community Child Health 2012, p. 5). Place-based approaches value community-specific definitions of health needs and solutions, as well as governance. They build community engagement and capacity to facilitate individual, systems and cultural changes that promote health and wellbeing in situationally and culturally relevant ways (Cummins et al. 2007); to do so, place-based approaches enlist the disadvantaged population to challenge the basis of their marginalisation—they make a location knowable and manageable through localised decision-making. PBIs therefore, represent a 'convergence between economic discourses, and discourses of inclusion, urban health and public health...within social policy' (Rushton 2014, p. 109).

Place-based health promotion programs can deliver significant dividends (Churchill et al. 2012), however, tensions have emerged regarding their implementation in the context of these projects. One of these tensions is between the organic responsive community development approach (which underpins place-based strategies), and the increasing use of 'rational' outcomes-oriented accountability frameworks by governments to demonstrate the efficacy of policies (Leviton et al. 2010; DHHS 2104). These outcomes frameworks typically require 'evidence' that interventions are having the desired impact on communities within relatively short timeframes. This 'evidence' includes the gathering and analysis of outcomes data, which community organisations are unlikely to collect routinely.

One solution to this issue is to create an 'evaluation culture' in community-based organisations through ECB activities, however, this is no simple task. Aside from the mechanics of designing and implementing a relevant and appropriate ECB program, there is a cultural dimension to consider. The core mission of community sector organisations is to respond to, support and provide services to their disadvantaged community (McCoy, Rose & Connolly 2013), rather than conduct evaluations. Further, the tools commonly used in program evaluation and outcomes frameworks (for example program logic models) can seem at odds with the way community development organisations work on the ground—often responding and reacting to changing needs as they arise. In the following discussion, we illustrate some of these challenges in the TCHF projects.

Building evaluation in the Thriving Communities Healthy Families pilot project

The TCHF pilot project included a funding component to build evaluation and planning capacity, in response to a perceived lack of an evaluation culture and evaluation skills within the NH network. Therefore, in



this context, ECB was primarily an opportunity for the Healthy Families workers and house coordinators to build new skills and knowledge. This would help them to demonstrate client outcomes and continuously reflect on and improve their programs. Further, this would prepare NHs to deal with the planned shift in contract arrangements between the community sector and the Tasmanian Government during 2014-15, that required funded community organisations to demonstrate program outcomes. NHT entered into an agreement with the University of Tasmania (UTAS) to evaluate the TCHF program and to deliver the ECB component.

An action learning approach was designed, whereby, evaluation training would be delivered in a series of workshops. As the project was implemented, participants would build evaluation capacity through 'learning by doing', drawing on the expertise of the UTAS evaluation team and NHT staff where needed. There were a number of assumptions underpinning the approach. Firstly, building evaluation capacity was a good and desirable thing; it was a professional development opportunity for the employees of the five houses to build new skills—for use in their current employment and future projects. Secondly, once armed with the relevant theories, tools and techniques, employees would be able to develop evaluation plans and design evaluation activities in their own projects. Third, once 'skilled up', employees would have the confidence to conduct evaluation activities and, with support from the evaluation team, use evaluation findings to inform program development and implementation. Finally, the TCHF program could be conceptualised as having a 'single' project framework within which the five individual projects would fit. Therefore, it was possible to use a core set of evaluation tools and techniques across all projects, despite their diversity.

Early into the project implementation, it became apparent that the envisaged ECB approach was unsuitable for three main reasons. First, and most importantly, each house was a unique organisation with a different history, skill set, area of expertise and community. In some houses, the house coordinator and Healthy Families worker were long established employees with strong personal connections to their communities and deep reserves of local knowledge. In other houses, the coordinator and Healthy Families worker were either new to the NH network, or were from outside the area and had to start building relationships with the community before they could build engagement strategies with at-risk families; this meant they had less time to engage in ECB as they focussed primarily on building trust and acceptance in the community. Secondly, the house coordinators and Healthy Families workers had a wide range of evaluation skills and knowledge, making it difficult to design training that was appropriate for the whole group. Finally, there was a significant logistical issue; participating houses were geographically dispersed and almost everyone worked

part-time, making it difficult to get people together for evaluation workshops for even half a day.

Consequently, ECB took place during intensive, individual three-hour sessions with each of the houses, the UTAS evaluator and the NHT project manager, at around six months after project commencement. Over the ensuing six months, the UTAS evaluator and the NHT project manager continued to work with each house separately to develop evaluation plans for the house-specific projects. The evaluation plans were originally earmarked as an early program deliverable that would help guide ongoing ECB needs however, they were still a work in progress 12 months after project commencement. Therefore, the management of the evaluation became an ongoing task, competing for the finite time the Healthy Families workers had for program design and implementation. The planning of the evaluation was also a much more resource-intensive process for the UTAS evaluator and project manager than first anticipated, which had budget implications.

As well as being an ECB strategy, the individual sessions were an opportunity to assess project progress from an overall evaluation perspective. The aim was to introduce tools and approaches to project planning and evaluation, whilst also identifying evaluation questions and potential performance measures. The sessions were discursive; they began with the evaluator asking the houses about the population they were going to work with, and the changes they wanted to see in that population as a result of the TCHF pilot project. From here, using a program logic approach, project staff identified potential engagement strategies and health promotion interventions/programs for their particular target population. They also identified potential key partners to work with and possible governance processes to manage the project.

At the project mid-point, the evaluation team conducted a series of semi-structured key informant interviews with Healthy Families workers, NH coordinators and project management staff. These interviews revealed a disconnect between the ECB approach taken and the needs of the participants. When the evaluation team introduced standard evaluation tools such as logic models, theory of change models, and the concept of an evaluation plan, it caused varying levels of anxiety amongst houses. It was clear that some houses found it problematic to articulate their projects in terms of target populations, outputs and outcomes. They perceived the program logic approach as static and somehow locking them into a particular path—this went against their usual responsive and dynamic way of working with their community. As one key informant commented:

[NH] are used to being responsive and each month they will hear new suggestions...it's a very reactive process...trying new things over the course of a year. [NH] are used to giving the community what they ask for, noticing if it's helpful and if it's not they will move



on to something else. [Its] not always a formal planning and evaluation process to test whether they achieved the outcome they were aiming for in the start (TCHF Key Informant Interview 2014).

Key informants also raised the issue of the diversity of evaluation skills and experiences across the participating houses. One informant noted ‘so much depends on the evaluation skillset of each house which are very different’. Another informant said ‘the starting point of the different players was quite diverse which makes designing anything quite difficult when it comes to building [evaluation] capacity’.

It was difficult for some houses to differentiate conceptually between evaluation planning and evaluation activities. It was problematic for them to see evaluation as a reflective process that begins with program planning and continues throughout program implementation, rather than being a singular ‘event’. For example, one key informant said ‘we feel we are evaluating before we’ve had time to implement’ and the evaluation was ‘so early into the programs’ (TCHF Key Informant Interviews 2014). The key informant saw evaluation as something that happens after a program is completed, rather than a continuous parallel process to program implementation. This highlights an important contextual consideration that before embarking on ECB, it is important that organisational norms and practices, as well as participants’ perceptions and assumptions about evaluation and its practice are understood (Preskill & Boyle 2008). This contextual consideration was never fully explored in the planning phase of the TCHF project, so the opportunity to design an ECB strategy that specifically met the range of needs for the NHs and project staff was not realised.

Reflecting on evaluation capacity building

Overall, TCHF project staff perceived evaluation and building evaluation capacity as worthwhile and valuable. They saw clear benefits to having the skills to build evidence and demonstrate the outcomes of programs. However, there were some conflicts around who should be doing evaluation and whether or not building evaluation capacity in frontline service delivery personnel was an appropriate or useful enterprise. One key informant said:

Evaluation is great but it’s someone else’s job and the service delivery people can’t do both...it is important to evaluate how programs are being delivered and utilised and what effects they have on the community, however, asking a service delivery person to become an evaluator is not only unrealistic, it is a drain on resources and it’s unfair to laden a service delivery person with a lot of evaluation training or responsibility’ (TCHF Key Informant Interviews 2014).

Other workers viewed the ECB more favourably, describing it as ‘a learning opportunity to develop evaluation skills’ and ‘once you get your head around it, it’s quite useful—if it can help us demonstrate what we do [and] secure more funds to support people then it’s worthwhile’ (TCHF Key Informant Interviews 2014).

For most, the time commitment associated with both ECB and evaluation activities was burdensome. There was a sense that it distracted the Health Families workers from the real work of supporting vulnerable clients. One interviewee said, ‘the evaluation aspect has been the hardest bit of my role. I only have 19 hours a week across 4 days to deliver the program, so evaluation is the hardest bit of that’ (TCHF Key Informant Interviews 2014). Most respondents commented that, given the short timeframe of the project, the relatively small amount of funding per house and the part-time appointments of Healthy Families workers—evaluation expectations were too high. This illustrates a structural tension in the current model of delivery of human services via the community sector. Organisations often receive funding that can only support part-time and contract work, at the same time, there is an expectation of more robust and sophisticated evaluation of individual projects, as well as the development of an embedded evaluation culture and capacity.

The Healthy Families workers (supported by their house coordinators to different degrees), were simultaneously designing the project, engaging with their clients, creating partnerships with other organisations and delivering health promotion interventions, leaving little time to engage positively with ECB. As one interviewee commented, ‘there is value in training and evaluation, the point is that if it is going to be so time-consuming, then additional resources need to be added onto the funding package’ (TCHF Key Informant Interviews 2014).

On reflection, the ECB component of the TCHF project was never fully articulated. As a result, there was no shared understanding between project managers, project staff and the evaluation team regarding the specific purpose or aims of ECB in the context of this project; there was also no mutual agreement on the approach to ECB that would best meet the needs of participants. As one key informant said:

The rationale behind developing a specific type of evaluation capacity was not clearly articulated and how this translates into building general evaluation capacity in this setting is not clear (TCHF Key Informant Interview 2014).

There was an assumption that offering evaluation capacity as a staff development opportunity was self-evidently of benefit to individual participants and as a result, an evaluation culture would begin to form within the NHs. This would better position the houses to demonstrate their program outcomes, respond to new accountability demands and secure continuing funding in a resource-



constrained environment. Further, there was an assumption that the Healthy Families workers (once introduced to relevant evaluation concepts, tools and methods), would engage with evaluation as a positive endeavour. The reality was somewhat different—ultimately, the ‘technical’ approach to building ECB in front line workers did not suit the needs of the TCHF projects. The short timeframe and the limited resources available for ECB did not allow a strong trust-based relationship to form between the evaluation team and the project workers. Evaluation was not a positive continuous reflective practice that enhanced program design and implementation, but rather another demand on already stretched project resources.

Conclusion

The TCHF experience gives pause for critical reflection on how best to support micro-community organisations to negotiate the current trend for evidence-based, outcomes-driven practice, for the enhanced accountability of public funding. These organisations exist for their community and are ‘of their community’, giving them privileged access and positioning them well to design and implement services for disengaged populations. Their primary focus will always be supporting their communities, but demonstrating the efficacy and impact of their programs is a reality in a competitive funding environment. Building evaluation capacity and instilling an evaluation culture in community organisations can help organisations reorient towards an outcomes focus and be better placed to meet increased accountability demands.

The ECB activities conducted with the TCHF project staff exposed the need for a bespoke approach. Training or professional development on its own does not constitute ECB (Beere 2008; Adams & Dickenson 2011). Training can improve evaluation capability of individuals, but is unlikely to improve evaluation capacity of organisations (Adams & Dickenson 2011) or indeed, create an evaluation culture (Beere 2005). Building evaluation capacity is essentially a discursive process; the tools, techniques and theories are a necessary starting point, but it is through practice and reflection that evaluation becomes integral to how an organisation operates. McCoy, Rose and Connolly say there is:

A compelling case for organisational investment in the creation of an organisational evaluation culture where staff value evaluation and are motivated to participate in demonstrating the difference that their work makes at an organisational and sector level (2013, p. 19).

The TCHF experience suggests this is unlikely to happen when the focus of the ECB is on frontline workers already juggling competing demands in a time-constrained and resource-poor environment. Under these circumstances, ECB was an additional burden that could demonstrate few immediate benefits. NH employees needed to build a partnership with the evaluation team

prior to embarking on the ECB journey, however, limited time and resources did not allow for this, especially when the TCHF project was fragmented into five separate programs. A partnership would have allowed for a mutually negotiated ECB approach, for trust to build, and, for participants to feel supported to apply evaluation skills and knowledge to their practice.

It seems clear that building evaluation capacity through ‘contract’ training as a component of specific project funding did not work well in this project. Perhaps something more akin to an ongoing mentorship model between community organisations and evaluation ‘expertise’ would be a better option for future ECB endeavours. For example, an NFP organisation could collaborate with a local university department to negotiate mutual benefits, such as continuing evaluation support in return for research outputs for the university. Further, it may be more beneficial to target ECB at the organisational management level rather than front line workers, if cultural change is the objective. Alternatively, ECB could target community sector peak bodies that could consolidate evaluation expertise and work in a collaborative model to support their member organisations.

Creating an evaluation culture requires investment by all parties. Funders placing increasing accountability demands on NFPs need to commit sufficient resources to building evaluation capacity in the system. NFPs need to embrace evaluation capacity as an asset that can help them meet external accountability demands; build the human services evidence base; and deliver better outcomes for their own clients. Small organisations need support and a realistic timeframe to create a culture that perceives evaluation and ECB as a benefit, not a burden. Further, it may be timely to rethink what evaluation capacity means for very small organisations in the community sector. Perhaps Bourgeois and Cousins’ (2013) concept of evaluation literacy is what is really needed—the capacity to understand and use evaluation, not necessarily the capacity to do evaluation. This gives cause for reflection on the unique roles that each component of the community sector system should play in both the building of sector-wide evaluation capacity and responding to the accountability challenge.

References

- Adams, J & Dickinson, P 2011, ‘Evaluation training to build capability in the community and public health workforce’, *American Journal of Evaluation*, vol. 31. no. 3, pp. 421–433.
- Australian Bureau of Statistics 2012, *Australian health survey: first results, 2011–12*, ABS catalogue no. 4364.0.55.001.
- Ernst & Young, 2014, *Research into Commonwealth regulatory and reporting burdens on the charity sector*, Australian Charities and Not-for-profits Commission, Canberra, viewed 24 February 2015, http://www.acnc.gov.au/ACNC/Pblctns/Rpts/EY_Report/ACNC/Publications/Reports/EY_report.aspx.



- ARACY 2009, *Measuring the outcomes of community organisations*, Australian Research Alliance for Children and Youth, viewed 17 February 2015, http://www.aracy.org.au/publications-resources/command/download_file/id/111/filename/Measuring_the_outcomes_of_community_organisations.pdf.
- Bacchi, C 2009, *What's the problem represented to be*, Pearson Education, Sydney.
- Banks, G 2009, 'Evidence-based policy-making: What is it? How do we get it?', paper presented at ANZSOG/ANU Public Lecture Series 2009, Canberra, viewed 23 February 2015, www.pc.gov.au/_data/assets/pdf_file/0003/85836/cs20090204.pdf.
- Beere, D 2005, 'Evaluation capacity-building: a tale of value-adding', *Evaluation Journal of Australia*, vol. 5, no. 5, pp. 41–47.
- Bourgeois, I & Cousins, JB 2013, 'Understanding dimensions of organizational evaluation capacity', *American Journal of Evaluation*, vol. 34, no. 3, pp. 299–319.
- Boyd, M 2008, 'People, places, processes: reducing health inequalities through balanced health promotion approaches', paper presented at the Victorian Health Promotion Foundation's People, Places, Processes Forum, April 2008, <https://www.vichealth.vic.gov.au/search/people-places-processes-forum>.
- Churchill, B, Doherty, T, Hansen, E & Eccleston, R 2012, 'People and place: understanding a research program for understanding and addressing place-based health inequities in Tasmania', literature review prepared for Department of Health and Human Services, University of Tasmania, <http://ecite.utas.edu.au/78536>.
- Cummins, S, Curtis, S, Diez-Roux, A & Macintyre, S 2007, 'Understanding and representing "place" in health research: a relational approach', *Social Sciences & Medicine*, vol. 65, no. 9, pp. 1825–1838.
- Dahlgren, G & Whitehead, M 1991, *Policies and strategies to promote social equity in health*, Institute of Future Studies, Stockholm.
- DHHS 2014, *DHHS funded Community Sector Outcomes Purchasing Framework*, Community Sector Relations Unit, Tasmanian Department of Health and Human Service, Hobart, viewed 17 February 2015, http://www.atdc.org.au/pdf/DHHS%20Funded%20Community%20Sector%20Outcomes%20Purchasing%20Framework_Final.pdf.
- Edwards, G 2012, 'Measuring and reporting nonprofit outcomes', *Interbusiness Issues*, June, viewed 17 February 2015, <http://www.peoriamagazines.com/ibi/2012/jun/measuring-and-reporting-nonprofit-outcomes>.
- García-Iriarte E, Suarez-Balcazar Y, Taylor-Ritzler T & Luna, M 2011, 'A catalyst-for-change approach to evaluation capacity building', *American Journal of Evaluation*, vol. 32, no. 2, pp. 168–182.
- Hoole, E & Patterson, TE 2008, 'Voices from the field: evaluation as part of a learning culture', in JG Carman & KA Fredericks (eds.), *Nonprofits and evaluation* [Special Issue], *New Directions for Evaluation*, no. 119, pp. 93–113. DOI:10.1002/ev.270.
- Huffman, D, Thomas, K & Lawrenz, F 2008, 'A collaborative immersion approach to evaluation capacity building', *American Journal of Evaluation*, vol. 29, no. 3, p. 358–368.
- Leviton, LC, Khan, LK, Rog, D, Dawkins, N & Cotton, D 2010, 'Evaluability assessment to improve public health policies, programs and practices', *Annual Review of Public Health*, vol. 31, pp. 213–233.
- Marston, G & Watts, R 2003, 'Tampering with the evidence: a critical appraisal of evidence-based policy-making', *Australian Review of Public Affairs*, vol. 3, no. 3, p. 143–163.
- McCoy, A, Rose, D & Connolly M 2013, 'Developing evaluation cultures in human service organisations', *Evaluation Journal of Australasia*, vol. 13, no. 1, pp. 15–20.
- Neighbourhood House Tasmania 2015, *What is NHT*, <http://nht.org.au/home/neighbourhood-houses-in-tasmania/>.
- Preskill, H & Boyle, S 2008, 'A multidisciplinary model of evaluation capacity building', *American Journal of Evaluation*, vol. 29, no. 4, pp. 443–459.
- Productivity Commission 2010, *Contribution of the not-for-profit sector*, research report, Canberra.
- Rushton, C 2014, 'Whose place is it anyway? Representational politics in a place-based health initiative', *Health & Place*, vol. 26, pp. 100–109.
- Solesbury, W 2001, *Evidence based policy: whence it came and where it's going*, working paper 1, ESRC UK Centre for Evidence Based Policy and Practice, Queen Mary University of London, London, viewed 23 February 2015, http://www.lgsp.uz/old/publications/option_paper_training/ebp_when_it_came_and_where_it_is_going_eng.pdf.
- Stockdill, SH, Baizerman, M & Compton, DW 2002, 'Toward a definition of the ECB process: a conversation with the ECB literature' in *The art, craft, and science of evaluation capacity building* [Special Issue], *New Directions for Evaluation*, spring 2002, vol. 2002, no. 93, pp. 1–26.
- Suarez-Balcazar, Y & Taylor-Ritzler, T 2013, 'Moving from science to practice in evaluation capacity building', *American Journal of Evaluation*, vol. 35, no. 1, pp. 95–99.
- TasCOSS 2014, *Measuring and reporting outcomes of community sector grants; a plain English resource to use with the DHHS Community Sector Outcomes Purchasing Framework*, Tasmanian Council of Social Services, Battery Point, Tasmania, viewed 17 February 2015, <http://tinyurl.com/q8ufne5>.
- The Royal Children's Hospital Centre for Community Child Health 2012, *Place-based initiatives transforming communities: proceedings from the Place-based Approaches Roundtable*, The Royal Children's Hospital Centre for Community Child Health, Melbourne, http://www.rch.org.au/uploadedFiles/Main/Content/ccch/CCCH_Place-based_initiatives_report.pdf.

**KELLY RIJSWIJK | DENISE BEWSELL | BRUCE SMALL | PAULA BLACKETT**

Reflexive monitoring in New Zealand: evaluation lessons in supporting transformative change

An Agricultural Innovation System (AIS) approach is currently being used to determine whether the adoption of agricultural innovations in New Zealand can be enhanced. Important elements of the innovation systems approach include the participatory application of a co-innovation process, for example inclusiveness and a focus on transformative change, to remove barriers in the creation and uptake of sustainable innovations. Alongside AIS, novel evaluation approaches are being used, namely reflexive monitoring. Thus, the role of a reflexive monitor has been trialled to help understand and guide the co-innovation process. A reflexive monitor helps innovation networks challenge and change assumptions, practices, and underlying institutions in the design or management of a project, through an on-going process of evaluation in an action research setting. In this article, experiences of reflexive monitoring conducted within a research programme in New Zealand are shared. Information was gathered from workshops, meetings and interviews with reflexive monitors. Insights include the need to understand that there is no recipe for the role of a reflexive monitor, that is the role is context-dependent. However, focussing on developing solutions to a problem and taking a flexible, adaptive approach are essential elements in ensuring that monitoring and evaluation are effective.

Introduction

As part of its Business Growth agenda, the New Zealand Government aims to increase innovation in business (MBIE 2014). The primary sector is the largest and most important sector of New Zealand's economy, contributing \$29.2 billion (Ministry of Primary Industries 2012) to New Zealand's \$206.5 billion GDP (Statistics New Zealand 2012). New Zealand's agricultural sector has remained internationally competitive, partly due to technological innovation and sensitivity to market changes and opportunities (Easton 2013).

However, much available science and technological innovation is not adopted by New Zealand farmers, primarily because it is not disseminated in a form that is usable or that can be readily adopted, from a 'whole-farm system' perspective (Fraser, cited in Chalmers 2014). Additionally, as social, economic and environmental

regulatory drivers require increasing consideration, the development and adoption of agricultural innovation becomes increasingly challenging. In this complex operating environment, where the conflicting needs and values of a range of different stakeholder groups need to be considered, the traditional technology transfer approach is failing to deliver the desired uptake of innovations (Botha et al. 2014).

Novel approaches and processes are required in order to address these challenges. Primary Innovation, a five-year research programme funded by the Ministry of Business, Innovation and Employment, is researching, testing and implementing a novel approach to address complex problems. This programme started in October 2012, and is designed to test the use of an Agricultural Innovation System (AIS) approach to enhance the usability, acceptability and adoption of agricultural



Kelly Rijswijk is a Junior Researcher in the Knowledge, Technology and Innovation Group, Wageningen University and Research Centre, Wageningen, The Netherlands. (Formerly a Scientist at AgResearch Limited, Ruakura Research Centre, New Zealand).
Email: Kelly.rijswijk@wur.nl



Denise Bewsell is a Researcher at AgResearch Limited, Lincoln Research Centre, New Zealand.
Email: denise.bewsell@agresearch.co.nz



Bruce Small is a Senior Scientist at AgResearch Limited, Ruakura Research Centre, New Zealand.
Email: bruce.small@agresearch.co.nz



Paula Blackett is a Social Scientist at AgResearch Limited, Ruakura Research Centre, New Zealand.
Email: paula.blackett@agresearch.co.nz

innovations in New Zealand. An AIS is a network of organisations, enterprises and individuals focussed on bringing new products, new processes, and new forms of organisation into economic use (World Bank 2006). The AIS approach uses a co-innovation process to address industry-focussed problems through a network of relevant stakeholders who are collaborating to develop new technologies or practices (Botha et al. 2014). According to Gilbert, Pyka and Ahrweiler (2001), 'Innovation is increasingly recognised as requiring the convergence of many sources of knowledge and skill, usually in the form of a network'. The AIS approach emphasises that agricultural innovation is not just about new technologies; it is a co-evolutionary process that may also require institutional change (Kilelu, Klerkx & Leeuwis 2013).

To monitor and evaluate progress towards sustainable solutions, a flexible approach is needed; this approach allows for ongoing adaptation within an ever-changing environment and enables the involvement of a wide range of stakeholders. Reflexivity has become an important part of a range of projects around the world, particularly where system innovation is required (Rogers 2008; van Mierlo, Arkesteijn & Leeuwis 2010a; Bussels, Happaerts & Bruyninckx 2013). Reflection, and co-reflection amongst participants, is advocated as an important process for the success of trans-disciplinary projects (Harris & Lyon 2010; Roux et al. 2010). Van Mierlo, Arkesteijn and Leeuwis (2010a, p. 145) highlight the need for system innovation projects to 'be accompanied by a monitoring and evaluation approach that supports and maintains such reflexivity'. Van Mierlo, Arkesteijn and Leeuwis (2010a, p. 145–146) emphasise the need for reflexive monitoring and evaluation to 'support complex projects without predefined outcomes...facilitate learning by questioning participants' values, and practices...and contribute to collective, institutional change, within and via the project'. This means that reflexive evaluation is not focused on learning as the end



point. Reflexive monitoring offers evaluators’ flexibility, an ability to respond to emergent trends rather than focusing on expected results, and a means of promoting participation. Reflexive monitoring and evaluation follows a cycle of observation, analysis, reflection and adjustment of project activities (van Mierlo et al. 2010b). Arkesteijn, van Mierlo and Leeuwis (2015) highlight that:

A reflexive perspective in evaluation...seek[s] to contribute to system change in order to deal with complex problems...[and] it should encourage groups of diverse actors to reflect on the rules and relations underlying current practices in order to induce institutional change (p. 108).

A practical means of achieving this is to nominate a member of the project team, or someone from outside the team to be the reflexive monitor, as outlined by van Mierlo et al. (2010b).

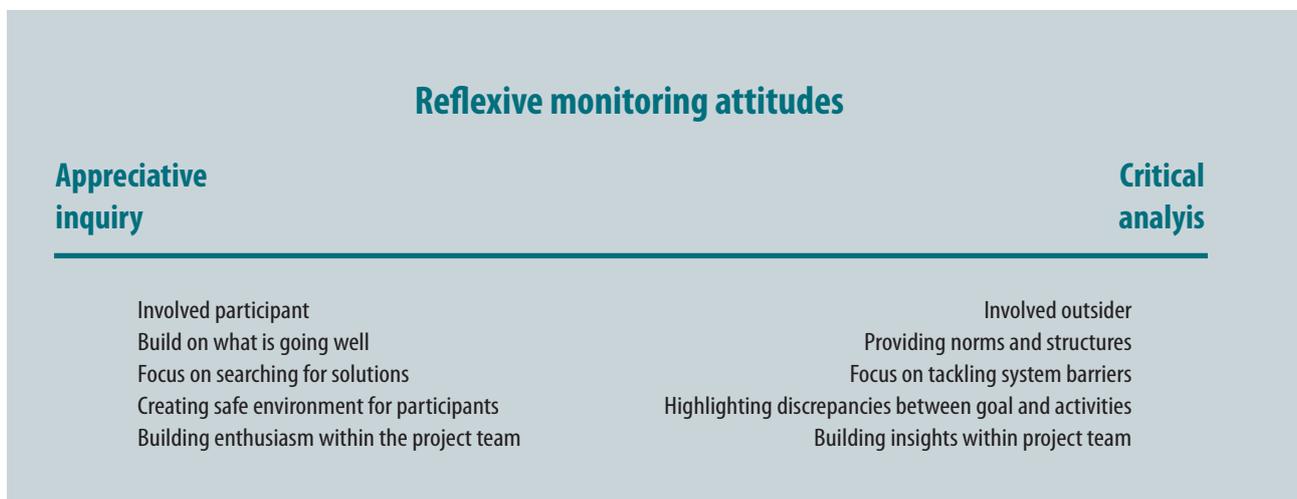
A reflexive monitor is a person who (through an ongoing process of evaluation in an action research setting) helps innovation project teams to reflect on process, action and progress towards the research goal, addressing barriers to both group dynamics and the AIS. The role may involve challenging and changing presumptions, current practices, and underlying institutions—either in the design or management of a project (Arkesteijn, van Mierlo & Leeuwis 2015; van Mierlo, Arkesteijn & Leeuwis 2010a; van Mierlo et al. 2010b). The role, therefore, involves monitoring and evaluation, but goes further, in that the reflexive monitor’s role is to ensure social learning is captured and can contribute to the transformational change required from the project. Van Mierlo et al. (2010b) outline two

ends of a spectrum for reflexive monitors: appreciative inquiry and critical analysis (see Figure 1). A reflexive monitor who operates at the appreciative inquiry end is an ‘involved participant’ in the project who helps to build enthusiasm within the project team. Whereas those operating at the critical analysis end are considered ‘involved outsiders’ who help to build insights within the project team. Reflexive monitors can operate anywhere along this spectrum; the situation or context will influence where a reflexive monitor operates along the system, and this may change throughout the life of a project.

Reflexive monitoring is used within Primary Innovation in five separate case studies or innovation projects (also known as innovation platforms), which address issues of differing complexity. These innovation projects are aiming to improve: i) dairy heifer rearing; ii) control of tomato-potato psyllid in potato crops; iii) water-use efficiency in an irrigation scheme; iv) implementation of nutrient management plans on dairy farms; and, v) the link between products from planted forests and markets. These innovation projects are trans-disciplinary and involve the integration of knowledge from multiple scientific disciplines, as well as local knowledge from a range of stakeholders to address industry problems. In such endeavours, group dynamics, and the relationships between project team members, is essential to project success (Burton et al. 2008).

In order to help remove barriers and enhance the success of the innovation projects, each project in the Primary Innovation programme has a reflexive monitor to help guide the co-innovation process. The reflexive monitors make use of nine principles that underpin and guide the co-innovation process. These nine principles

FIGURE 1: SPECTRUM OF ATTITUDES IN APPROACHING THE ROLE OF A REFLEXIVE MONITOR



Adapted from van Mierlo et al. 2010b



are: i) take time to understand the problem from many different views; ii) be inclusive; iii) engage with and value all sources of knowledge; iv) strive to learn from each other by actively listening and understanding; v) keep sight of the shared vision or 'ambition for change'; vi) be honest, open and constructive in interactions with other participants; vii) be aware of the wider context of the problem and any actual or potential changes which may occur; viii) be flexible and adaptable; and ix) adhere to the co-innovation process despite its frustrations (adapted from Nederlof, Wongtschowski & van der Lee 2011). The process used in the innovation projects, as well as participants' behaviour, is monitored against these principles on an ongoing basis to assess if the projects are meeting the end goals and needs of participants. The research team and the reflexive monitors in Primary Innovation also found Wiczorek and Hekkert's (2012) 'systemic innovation policy framework' as a useful tool for analysing innovation systems and identifying barriers; it also provides systemic tools to address barriers identified.

Methods

Data was collected on the experiences of the reflexive monitors' in each innovation project within the Primary Innovation programme. This data was gathered through discussion and feedback from monthly phone meetings and three face-to-face reflexive monitor meetings. Additionally, some reflexive monitors were interviewed as part of a wider programme evaluation in March 2014. An ORID (Stanfield 2008) evaluation of the individual innovation projects in July 2014 provided further input to this article, as well as the authors' experiences as reflexive monitors. These data were used to discern common themes in the reflexive monitoring experience which are discussed in the results section below. The experiences highlighted in the themes were used to reflect on insights into the practice of evaluation, given the experience to-date with reflexive monitoring in New Zealand.

Results

Although participatory research and adaptive management are research techniques used in New Zealand, the AIS approach and reflexive monitoring roles are new. Even though some guidance is available in scholarly literature, such as the management of trans-disciplinary research and innovation system theory, time was required for the Primary Innovation research team to understand the function and practice of reflexive monitoring, and how this could be linked to evaluation. A number of lessons have been learned about the function and practice of reflexive monitoring within the innovation projects and the wider Primary Innovation programme, as well as the specifics of the role of a reflexive monitor and the practice of evaluation.

Understanding theory

A research team workshop held in August 2013 (nearly a year after the Primary Innovation programme began), revealed that researchers (including the reflexive monitors), struggled to understand the theory of AIS, the systemic innovation policy framework (Wiczorek & Hekkert 2012), and the role of reflexive monitoring. Key questions for the research team were:

- (i) how much understanding of AIS theory is required to function effectively as a reflexive monitor?
- (ii) what other theories can inform how reflexive monitors act in practice?

Several reasons were identified regarding the struggles faced by the reflexive monitors. Firstly, most of the innovation projects started slowly, with considerable negotiation required to organise contracts between the participating research and business organisations. Secondly, several researchers involved in Primary Innovation (including some reflexive monitors) did not have a social sciences background. This hampered progress in understanding reflexive monitoring due to a lack of familiarity with participatory and adaptive management approaches. Initially, reflexive monitoring seemed too abstract and the level of theoretical knowledge of AIS required (before being able to undertake the role of reflexive monitor) was uncertain.

New role development

Although the operationalisation of AIS theory appears to require a reflexive monitor, there is little guidance on practically how this role can be performed, with the exception of van Mierlo et al. (2010b). Given the recent introduction of the role to this organisation in New Zealand, monitors were unsure what actions to take. This occurred even when monitors conceptually understood the reflexive monitoring approach and its function within Primary Innovation. Questions regarding the role of the reflexive monitor included:

- (i) would it be a facilitator role, an observer role or a bit of both?
- (ii) what would their level of participation in the innovation project be?
- (iii) what would a reflexive monitor observe?
- (iv) what are their responsibilities?
- (v) how would they address any issues they discovered?
- (vi) what tools were available to them?

One reflexive monitor commented on the need to be working with the project manager 'to define the role and what their expectations were'.

The lack of clarity regarding the practicalities of the role was also evident in the wider research team, and not



just those assigned to be reflexive monitors; this means that working with a project leader to define the role and expectations of the reflexive monitor was difficult for some. Although job descriptions were drafted for the position, a lack of experience with participatory approaches and related skills posed a barrier to confidence in undertaking the role for some reflexive monitors. Moreover, the role of a reflexive monitor, and the actions that a reflexive monitor can take, largely depend on the context in which reflexive monitoring takes place. Contextual factors might include: the number of stakeholders involved in the innovation project; the topic of the project; the level of complexity of the problem; whether the reflexive monitor works for the same organisation as the project team; and whether the monitor was proximally located (for ease of meeting) to the project leader and other key participants in the innovation project. These factors influence the level of understanding and skills required of the reflexive monitors. As a result, a set of initial guidelines and supporting questions were developed for the Heifer Rearing Innovation Project. The reflexive monitor was able to use these guidelines as a tool to explain the role and demonstrate outcomes from the monitoring undertaken.

Skills required

Based on our experiences, reflexive monitors need to be action-oriented persons, who are good at observing and listening, and continuously assess the situation of the innovation project. The monitors need the ability to make rapid decisions to act if the projects and processes do not meet the requirements of co-innovation; this requires facilitation skills and, if the situation involves conflict, mediation skills. Projects had to start according to contract dates, despite initially high levels of uncertainty about the theory and approaches used, and the role of reflexive monitors in the innovation projects. Reflexive monitors had to be flexible in dealing with uncertainties, proactive in making decisions, and inventive in testing ideas and taking action; these requirements tested the skills of all reflexive monitors and highlighted the need for ongoing training in this area. One reflexive monitor commented that an individual in the role ‘must have the skills to “speak the truth kindly” and remain dispassionate when those who are personally involved get defensive when you touch a nerve’.

Need for tools

Initially, reflexive monitors used a manual developed by van Mierlo et al. (2010b) to start activity in the innovation projects. Although this has been a good place to begin, it has become clear that the approaches will need to be extended and modified to suit the New Zealand context. This has led to the reflexive monitor team starting to develop their own guide to monitoring, based on the tools used in the innovation projects.

Trust

The co-innovation process requires considerable time and trust amongst the involved participant stakeholders. The role of a reflexive monitor is to help build this trust and create the space for equal inputs by participants, including the valuing of each other’s knowledge. Building relationships with mutual respect and trust is key to success. A reflexive monitor highlighted the need to ‘always be willing to see another point of view, and encourage others to see other points of view also’. In addition, participants and funders in such projects need to be aware that this is a time consuming and resource-intensive process. The focus is on encouraging full and equal participation to thoroughly understand the problem from all participants’ perspectives, before developing and agreeing upon the solution.

Discussion

The literature suggests that reflexive monitoring is most suited to situations where there is a high need for structural change, and an equally high need for input from all stakeholders to develop a solution for the required system changes (van Mierlo, Arkesteijn & Leeuwis 2010a). This means that reflexive monitors take an action research approach, incorporating evaluation into their role, and using social learning as one of the mechanisms to meet the goal of developing solutions to complex problems together. The reflexive monitor is there to help keep the project focused on this outcome, regardless of the barriers or contextual influences the project might encounter. In effect, the reflexive monitor is evaluating progress towards the required change and highlighting the need for adjustment.

The results outlined above from on-the-ground reflexive monitoring practice illustrate the difficulties involved in undertaking reflexive monitoring. Besides overcoming the above difficulties to help develop the innovation projects, the reflexive monitors needed to acquire legitimacy for their role in the project team. This was achieved by creating a range of opportunities for exploration and interaction with (potential) participants. Both time and encouragement were required by the project team in order to enable reflexive monitors to become more confident in their practice and influence, and to adapt to the specificities of each innovation project. The innovation project participants also had to understand that a co-innovation process requires more of their time and effort at the start of a project, but that it benefits long-term outcomes. Van Mierlo et al. (2010b) emphasise the need for flexibility in the practice of reflexive monitoring, this means working through the issues outlined above and working with a level of uncertainty. While there are no official guidelines for the role of a reflexive monitor, the reflexive monitors have



started to develop tools and guidelines to support and clarify their role.

The role of reflexive monitor seems to encapsulate the idea of developmental evaluation (Patton 2006) and provides a practical means of achieving this type of evaluation. The absence of a fixed approach for evaluation seems to be part of a reflexive practice and evaluation in these contexts. Given this, it seems that the need to allow for the building of trust—not just amongst project participants, but also in the reflexive monitor role—is a key insight for future projects and this type of evaluation.

Acknowledgements

We would like to acknowledge the Ministry of Business, Innovation and Employment who are funding the Co-learning and Co-innovation for Increased Impact project (CONT-30071-BITR-AGR) through a Biological Industries Targeted Research grant. A special thank you goes to the other reflexive monitors in New Zealand—Nicola Park, Janet Reid, and Karen Bayne, for sharing their experiences and thoughts. A special thank you also to Tracy Payne for undertaking interviews with the reflexive monitors and capturing some of the lessons outlined here.

References

- Arkesteijn, M, van Mierlo, B & Leeuwis, C 2015, 'The need for reflexive evaluation approaches in development cooperation', *Evaluation*, vol. 21, no. 1, pp. 99–115.
- Botha, N, Klerkx, L, Small, B & Turner, J 2014, 'Using co-innovation to stimulate change in the New Zealand agricultural sector', *Agricultural Outlook*, vol. 43, no. 3, pp. 219–223.
- Bussels, M, Happaerts, S & Bruyninckx, H 2013, *Evaluating and monitoring transition initiatives. Lessons from a field scan. Research paper 5*, Policy Research Centre Transitions for Sustainable Development, Leuven, Belgium, <https://steunpunttrado.be/documenten/papers/trado-rp-5-evaluating-and-monitoring.pdf>.
- Burton, L, Westen, D & Kowalski, R 2008, *Psychology*, John Wiley & Sons, Milton, Queensland.
- Chalmers, H 2014, 'Scientists need to learn how to make research accessible', *The Press*, 11 July 2014, Christchurch.
- Easton, B 2013, *Economy-Agricultural production*, Te Ara-the Encyclopedia of New Zealand, viewed 14 January, <http://www.TeAra.govt.nz/en/economy/page-2>.
- Gilbert, N, Pyka, A & Ahrweiler, P 2001, 'Innovation networks—a simulation approach', *Journal of Artificial Societies and Social Simulation*, vol. 4, no. 3, <http://jasss.soc.surrey.ac.uk/4/3/8.html>.
- Harris, F & Lyon, F 2013, 'Transdisciplinary environmental research: building trust across professional cultures', *Environmental Science and Policy*, vol. 31, pp. 109–119.
- Kilelu, CWL, Klerkx, L & Leeuwis, C 2013, 'Unraveling the role of innovation platforms in supporting co-evolution of innovation: contributions and tensions in a smallholder dairy development programme', *Agricultural Systems*, vol. 118, pp. 65–77.
- MBIE 2014, *Business growth agenda*, Ministry of Business, Innovation and Employment, Wellington, viewed 14 January, <http://www.mbie.govt.nz/info-services/business/business-growth-agenda>.
- Ministry of Primary Industries 2012, 'Tough going for primary products', media release dated 12 December, Rural News Group 2012, Auckland, <http://www.ruralnewsgroup.co.nz/rural-news/rural-general-news/tough-going-for-primary-products>.
- Nederlof, S, Wongtschowski, M & van der Lee, F (eds) 2011, *Putting heads together. Agricultural innovation platforms in practice. Bulletin 396*, KIT Publishers, Amsterdam, The Netherlands.
- Patton, MQ 2006, Evaluation for the way we work, *The Nonprofit Quarterly*, vol. 13, no. 1, pp. 28–33.
- Rogers, PJ 2008, Using programme theory to evaluate complicated and complex aspects of interventions, *Evaluation*, vol. 14, no. 1, pp. 29–48.
- Roux, DJ, Stirzaker, RJ, Breen, CM, Lefroy, EC & Cresswell, HP 2010, 'Framework for participative reflection on the accomplishment of transdisciplinary research programs', *Environmental Science & Policy*, vol. 13, no. 8, pp. 733–741.
- Stanfield, B 2008, *The Art of Focused Conversation: 100 ways to Access Group Wisdom in the Workplace*, Canadian Institute for Cultural Affairs.
- Statistics New Zealand 2012, 'Size of New Zealand economy reaches \$207 billion', media release dated 21 November, http://www.stats.govt.nz/browse_for_stats/economic_indicators/NationalAccounts/NationalAccountsIncomeExpenditure_MRYeMar12.aspx.
- Van Mierlo, B, Arkesteijn, M & Leeuwis, C 2010a, 'Enhancing the reflexivity of system innovation projects with system analyses', *American Journal of Evaluation*, vol. 31, no. 2, pp. 143–161.
- Van Mierlo, B, Regeer, B, van Amstel, M, Arkesteijn, M, Beekman, V, Bunders, J, de Cock, T, Elzen, B, Hoes, AC & Leeuwis C 2010b, *Reflective monitoring in action. A guide for monitoring system innovation projects*, Communication and Innovation Studies, Wageningen University/Athena Institute, Amsterdam.
- Wieczorek, AJ & Hekkert, MP 2012, 'Systemic instruments for systemic innovation problems: a framework for policy makers and scholars', *Science and Public Policy*, vol. 39, pp. 74–87.
- World Bank 2006, *Agricultural innovation systems: an investment sourcebook*, World Bank, Washington DC, <http://siteresources.worldbank.org/INTARD/Resources/335807-1330620492317/8478371-1330712129614/9780821386842-corrected.pdf>.

Title: *Interviewing for Qualitative Inquiry: A Relational Approach*

The author of the book *Interviewing for Qualitative Inquiry: A Relational Approach* is Ruthellen Josselson, a Professor of Clinical Psychology at Fielding Graduate University in Santa Barbara, California. The author draws on her 20 years' experience in conducting workshops in the United States and abroad, to explore the intricacies of interviewing for research purposes.

Josselson writes in an engaging fashion and answers in detail those many questions arising from early career researchers who are embarking on in-depth interviews as part of their qualitative research. The author endeavours to gauge the ideal structure of 'the interview' in order to elicit a narrative that responds directly to the research question. The interview process is described in great detail—from the first moment of recruiting participants, through to the post-interview stage of transcription and report preparation. The text is never prescriptive, instead, the reader is drawn into the topic with persuasive gentleness.

Chapter one emphasises the importance of the research relationship between the interviewer and the interviewee. In particular, the chapter details how this relationship influences the process of eliciting the participant's life experience and importantly, the participant's understanding of the fundamental research question.

Chapter two explores the myth of the neutral interviewer. The author discusses various aspects involved in preparing for interviews, including the interviewer's own experiences and expectations, and their anxiety, ethical attitude and potential power imbalance with the interviewee. Josselson draws attention to the importance of the anonymity guarantee, of confidentiality and of the assurance that the interviewee can withdraw from the research study at any time.

Chapter three discusses the focus and open-ended design of the interview. Josselson skilfully outlines clear and concise steps for the interview process, with accompanying detailed discussion about the pitfalls that the interviewer may encounter, for example—the danger of allowing the interview to fall into a question-and-answer format. This chapter contains examples of conceptual questions that should be extremely useful to the novice interviewer, for example, the initial expression of the study as a general statement ('The Big Q' p. 36). The author then points out the importance of engaging the interviewee at the start of the interview through a point close to his or her experience that is, 'the little q' (p. 41). Interview aids are also discussed, with examples provided.

Chapter four examines the beginning of the interview; the choice of venue; the equipment and placing of microphones; examples of opening remarks; and, the process of obtaining informed consent from the interviewee. Also discussed in this chapter is the importance of the 'empathic response'; the value in the interviewer reflecting their understanding of the interviewee's story and following 'where the interviewee goes' (p. 78), the absence of this response is to risk disrupting the interview by interrupting and attempting to bring the conversation back to the original focus.

■ Author

Ruthellen Josselson

■ Publisher/year

Jossey-Bass, San Francisco, second edition, 2013

■ Extent/type

206 pages, paperback

■ Price

AU\$54.95/NZ\$55.65 from Footprint Books which offers a 15% discount to AES members; phone 1300 260 090 (within Australia), or +61 2 9997 3973 (from outside Australia). Email info@footprint.com.au, website <http://www.footprint.com.au> for the latest prices

■ ISBN 978-1-4625-1000-9



Reviewed by **Ruth Davey**

Project Officer
Community Accommodation and
Respite Agency (Cara)
Woodville, South Australia
Email: ruth.davey@internode.on.net



Chapter five explores the interviewee's worldview through the utilisation of the 'empathic attitude of listening' (p. 80); examples are given of the process of deeper inquiry to further expand the interviewee's narrative. The advantages of paraphrasing are also mentioned as a means of focusing on a particular line of inquiry.

Chapter six describes the role of ethics and humanity as integral to the research relationship. In particular, the author points out the danger of the interviewer applying his or her interpretation to the participant's contribution—she notes it is crucial that the interviewee remains in control of his or her story. Josselson emphasises the importance of recording observations (taken prior to, and during the interview) as soon as possible after the interviewee has left the venue. She describes the researcher's ethical duty to protect the identity of participants, places and the agencies associated with those participants in the treatment of transcripts and published material. This chapter concludes with sample remarks for closing an interview, along with valuable advice to request feedback from the interviewee on his or her interview experience.

Chapter seven gives the reader examples of 'good' interviews with accompanying discussion of 'common threads' (p. 132) contributing to their success. The author recommends reviewing the transcription of the interview with the interviewer's supervisor or peer group in order to identify where the interviewer succeeded in 'moving with' the interviewee and where the interviewer may have gone off track.

Chapter eight raises awareness of 'bad' interviewing techniques that should be kept in mind whilst preparing for the interview. Advice is also given to cope with difficult situations, for example, when encountering interviewees that seem garrulous, distressed, hostile or boring.

The 'Do's and don'ts of interviewing' that are covered in chapter nine build on the author's previous advice and techniques. The importance of placing the interview in context and identifying the antecedent prior to disclosure will assist the interviewer with his or her future interviewing technique.

Chapter ten addresses what happens after the interview, particularly, the importance of recording impressions, observations and insights as quickly as possible; transcription and analysis are also discussed. The author refers to the work of Lincoln and Guba

(1985) with regard to 'member checking' (p. 178), but does not advise implementing this process where participants are asked to comment on the accuracy of the interviewer's interpretation of the interview, unless there is a very good reason for doing so. The author argues that the researcher's interpretation is 'a truth' rather than 'the truth' (p. 178). Josselson remarks that 'asking for participants' input does not solve the problem of locating 'truth' and creates other problems (p. 179). In any case, the reviewer puts forward for consideration the work of Rice and Ezzy (1999), who advocate for a clear process to be communicated in deriving themes from the interview. An additional method is to engage independent analysts to identify important points in the transcripts; the researcher can then compare the analysts' findings with his or her interpretation of the transcript (Davey 2008).

The conclusion in chapter eleven highlights the value of the described method of interviewing, and further, the information obtained through various types of analyses described by the author—narrative, phenomenological, discourse or grounded theory. The appendices contain useful interview aids, sample additional questions and a sample informed consent form; the latter helpfully outlines in detail issues that the interviewee should be aware of before signing the form.

Overall, the theme of *practicality* is present throughout the book; exercises appear in a number of chapters to give the reader practice and appreciation of important points of discussion. The author's views are robust and cover all aspects of relational interviewing. In particular, attention is given to formulating the questions, recruiting the participants and choosing the venue. A checklist of equipment and forms is provided and this is essential reading. The author provides excellent advice on conducting a 'good' interview; the process of transcription and analysis; and the production of material for publication. I recommend this book—it will especially appeal to students, given the wealth of examples of good practice.

References

- Davey, VR 2008, 'Exploration of the relationship of self-employment and quality of life', PHD thesis, Flinders University, Adelaide.
- Rice, PL & Ezzy, D 1999, *Qualitative research methods: a health focus*, Oxford University Press, South Melbourne, Victoria.

Notes to contributors

AES members and people involved in the practice, study or teaching of evaluation are invited to submit papers, practice articles and book reviews for publication. Before submitting, contributors are advised to consult the full guidelines set out on the AES website: www.aes.asn.au.

Decisions regarding publication are made on the basis of assessment by the editors, with advice from expert reviewers in a relevant field. Submissions to the EJA are sent for formal peer review and refereed articles are identified accordingly. However, the editors reserve the right to publish an article as refereed or non-refereed as appropriate.

There is no predetermined 'house style' for EJA. Authors are encouraged to give expression to their own individual voice, provided it meets generally accepted standards of clear, accurate and literate English usage. The principles set out in the *Style Manual for Authors, Editors and Printers* (sixth edition) should be used as a general guide to usage. Citations should follow the Harvard (i.e. author–date) style.

Authors are asked to write their articles in a way that foregrounds the most significant and interesting features of their content, avoiding unfamiliar technical language, extensive data tables and mathematical notation, except to the extent that these are necessary to convey the argument. Articles should be arranged in a way that facilitates comprehension by non-specialists, for example by using text boxes for background information and aspects of detail that are not central to the exposition.

References should be kept to the minimum required for understanding the text or following up sources specifically quoted or discussed in the article.

The editors reserve the right to correct obvious errors in spelling, vocabulary and grammar without reference to the author. Any further more extensive corrections or modifications that the editors and/or peer reviewers consider desirable will be discussed and explicitly cleared with the author before being incorporated in the published text.

Individual contributions should not normally exceed a published length of 10 pages, including all tables, graphics, notes and references. The editors reserve the right to edit longer articles down to this length if space considerations require. Articles shorter than 3000 words will not generally be regarded as suitable for peer review.

Copyright in articles submitted for publication in EJA resides with the authors, unless otherwise specified. Copyright in the form of the articles as presented in EJA resides with the AES.

Articles should be submitted through the online submission form available on the AES website: www.aes.asn.au under *Resources*.

Full guidelines for contributors are published on the AES website: www.aes.asn.au.

AES BOARD
Information at
www.aes.asn.au



EVALUATION
Journal of Australasia

Editors

Dr Lyn Alderman
Associate Director
Academic Quality and Standards
Learning and Teaching Unit, Chancellery
Queensland University of Technology
Brisbane

Associate Professor Janet Clinton
Director, Centre for Program Evaluation
Melbourne Graduate School of Education
The University of Melbourne

Contact:

publications@aes.asn.au

Book Review Editor

Dr Darrel N Caulley, Evaluation Consultant

Assistant Editors

Dr Bronwyn Rossingh, Charles Darwin University
Dr Carol Quadrelli, Queensland University of Technology
Dr Liz Gould, NSW Department of Education and Communities

Journal and Publications Working Group

- Lyn Alderman, Academic Quality and Standards, Learning and Teaching Unit, Chancellery, Queensland University of Technology, Queensland (Chair, *EJA* Editor)
- Sarah Appleton-Dyer, Synergia, Auckland
- Paul Burton, Urban Research Program, Griffith University, Gold Coast Campus, Queensland
- Janet Clinton, Centre for Program Evaluation, Melbourne Graduate School of Education, Melbourne, Victoria (*EJA* Editor)
- Rick Cummings, Educational Development Unit, Murdoch University, Western Australia
- Liz Gould, NSW Treasury, Sydney, NSW
- Graeme Harvey, Strategic Policy and Planning Division, Department of Education and Training, Victoria
- Ian Patrick, Ian Patrick & Associates, Victoria

Editorial Advisory Board

- Anona Armstrong, Victoria Law School, Victoria University, Victoria
- Darrel N Caulley, Evaluation Consultant, Victoria
- Rick Cummings, Educational Development Unit, Murdoch University, Western Australia
- Sue Funnell, Performance Improvement, NSW
- Penny Hawkins, The Rockefeller Foundation, New York
- Chris Milne, ARTD, NSW
- John Owen, Centre for Program Evaluation, The University of Melbourne, Victoria
- Colin Sharp, P.E.R.S.O.N.A.L. (Research and Evaluation) Consultancy Pty Ltd, South Australia
- Ralph Straton, Institute for Social Programme Evaluation, Murdoch University, Western Australia
- Ian Trotman, New Zealand
- Zita Unger, Ziman Pty Ltd, Victoria
- Yoland Wadsworth, Australia
- Jerome Winston, Program for Public Sector Evaluation International, Victoria

ISSN 1035-719X

The Australasian Evaluation Society (AES) is a professional organisation for people involved in evaluation. It has over 1000 members engaged in all aspects of evaluation and performance measurement. Members include evaluation practitioners, managers, teachers and students from all levels of government, educational institutions, research agencies, community organisations, businesses and interested individuals. Members meet regularly through Regional Groups in major cities in Australia and New Zealand.

The AES aims to improve the theory, practice and use of evaluation through: establishing and promoting ethics and standards in evaluation practice; providing a forum for the discussion of ideas including AES publications, seminars and conferences; linking members who have similar evaluation interests; providing education and training in matters related to evaluation; recognising outstanding contributions to the theory and/or practice of evaluation; acting as an advocate for evaluation in Australasia; and other activities consistent with its aims.

The purpose of the Evaluation Journal of Australasia is to further the aims of the AES by publishing articles of professional merit on any subject connected with evaluation, together with news and information of professional interest to members of the AES. Preference is given to articles that embody original concepts, significant new findings or methodological advances, or which present existing knowledge in a form particularly accessible or useful to students, practitioners and/or users of evaluation. The journal is intended to be critical and intellectually rigorous, but at the same time readable and accessible to the widest possible range of scholars and practitioners.

ISSN 1035-719X



Australasian Evaluation Society Ltd

PO Box 476, Carlton South VIC 3053, Australia

Tel: +61 3 9035 3469

Email aes@aes.asn.au

Website www.aes.asn.au