

Developing a thoughtful approach to evaluation: Values-driven guidelines for novice evaluators

Kelsey L. Deane and Niki Harré

Novice evaluators in Aotearoa New Zealand have few resources to assist them with reconciling the complex philosophical, theoretical, and practice tensions inherent in programme evaluation. This article provides guidelines to assist novice evaluators in their journey to become “thoughtful” practitioners balancing methodological credibility with stakeholder empowerment. The approach is informed by an analysis both of key themes and tensions identified in the evaluation literature, and of the first author’s experience as a novice evaluator contending with the difficulties of first articulating, and then advancing, the values she sought to embed in her practice.

Introduction

Programme evaluations are undertaken to achieve two overarching goals: (1) to gain knowledge about a programme to make a defensible judgement about its value, worth, or merit; and (2) to use that knowledge to inform decisions and future action (Scriven, 1967). Yet these goals are far from straightforward. Real-world evaluations are enacted

within a complex web of personal and societal values (Bamberger, Rugh, & Mabry, 2006; Greene, 2011; Kirkhart, 2015). These values impact on what is considered “knowledge”, how to gather that knowledge and how to respond to it. A further complexity is that much of the extant evaluation literature is dominated by philosophical, theoretical, and values-driven debates. This complexity may offer little practical assistance to the novice evaluator who, we suggest, needs rules of thumb that point to the core considerations inherent to any evaluation. As suggested by Nunns, Peace, and Witten (2015), many individuals practising evaluation are disconnected from the academic debates that provide a bridge between complex theories and practice. While initial and ongoing exposure to such debates is important, it is also extremely useful to provide guidelines that explicitly attempt to bridge that gap.

Here, we offer guidelines to the novice evaluator—what we call a *thoughtful* approach to evaluator development. As such, this article provides some stepping stones to move novices towards effective and reflective evaluation practice. We begin with an overview of the core issues and dilemmas we have identified in the evaluation literature that feed into the thoughtful evaluation approach. This includes a consideration of the particular issues relevant to Aotearoa New Zealand. We conclude with insights from the first author’s developmental journey as a *thoughtful evaluator*, highlighting practice tensions, successes, and lessons learnt along the way.

Evaluation: A debated discipline

The beginning: A focus on accountability, cost-effectiveness, and outcomes

Evaluation as a distinct field of practice gained substantial momentum during the 1960s. Key to the field’s proliferation was the surge

of government funding in the United States for social programmes during Lyndon B. Johnson's Great Society (Rossi & Freeman, 1989; Shadish, Cook, & Leviton, 1991). Programme funders soon realised that systematic processes were needed to determine which endeavours were legitimate and worth funding (Shadish et al., 1991). With social and behavioural scientists realising that research methods could be applied to gain knowledge of real-world problems, social-science methods began being used to determine the worth and merit of the different programmes (Rossi & Freeman, 1989).

Because the interest in evaluation during this early stage was focused on discerning the success or failure of programmes, evaluation questions were centred on whether programme objectives were being met (that is, outcome evaluation), how programmes fared in comparison to each other (comparative assessments), and which were cost-effective (cost–benefit analyses). Decision-makers wanted judgements that “summed up” the overall success of an initiative. This led to what is now known as summative evaluation (Scriven, 1967).

The social sciences, as with the natural sciences, at that time tended to rely on positivist worldviews, and methodologies primarily aligned with the quantitative-experimental paradigm (House, 1993; Mark, 2003; Patton, 2002). According to this view, the only way to obtain veritable knowledge (or “truth”) is through the measurement and manipulation of variables, and direct, objective observation of cause and effect (Trochim & Donnelly, 2008). When this paradigm was applied to evaluation, randomised controlled trials were (and, according to many, still are) seen as the gold standard because of the superiority they afforded for making causal claims about a programme's effects (Rossi & Freeman, 1989). Nevertheless, strict adherence to the quantitative-experimental paradigm as the sole approach to programme evaluation was criticised on many grounds (House, 1993; Patton, 1997; Shadish et al., 1991).

First, when evaluating mean differences between programme and control participants post-programme, null effects were obtained in the majority of cases (House, 1993). However, programme practitioners and evaluators who observed some of these programmes in action reported a different story (Patton, 2002). In essence, the “truth” about complex programmes was being reduced to performance on a small number of measurable indicators that failed to capture both the complex nature of programmes (Patton, 1997) and the differing views of stakeholders (House, 1993; Patton, 1997, 2002). This problem was exacerbated by evaluators who distanced themselves from stakeholders and programme operations (and thus a more rounded understanding of the programme) in order to maintain their objective stance. Information about the programme was communicated in scientific and statistical terms in reports that were often incomprehensible or irrelevant to stakeholders, further alienating them from the evaluation (Patton, 1997).

What is more, evaluators realised that implementing social experiments within real-world settings was much more difficult than in a research lab (House, 1993; Patton, 1997). Reported difficulties included sabotage of the implementation process by uncooperative practitioners (Weiss, 1998), participants receiving differential exposure to programme services, participant attrition, and control participants engaging with other programmes that offered similar services (Lipsey & Cordray, 2000; Weiss, 1998). All these factors affected counterfactual claims regarding programme effectiveness.

Additionally, evaluations that focused exclusively on programme effectiveness generated no insights about how or why a programme did or did not work (Chen, 1990; Patton, 1997). This created difficulties for those wanting to replicate the successful aspects of a programme or make improvements. Accordingly, they came to be known as “black box” approaches because people could not discern

what was occurring within the programme to produce or negate change (Chen, 1990). These factors all contributed to the evaluation “utilization crisis” (Patton, 1997, p. 7) that soon followed the initial growth in programme evaluation.

The utilization crisis: Increasing use through stakeholder involvement and attention to processes

The “utilization crisis” prompted a series of alternative approaches based on attention to programme processes and involving stakeholders in the evaluation¹ (House, 1993; Johnson et al., 2009; Mark, 2003; Patton, 1997; Weiss, 1998). These “process-focused” approaches were more informative for programme improvement and expansion and they became legitimate forms of evaluation (House, 1993; Patton, 1997; Weiss, 1998). A notable shift in researcher–stakeholder relations also eventuated as researchers began to recognise the importance of respecting and listening to those directly involved with programme initiatives. Related to these process-focused and participatory approaches, the qualitative-naturalistic paradigm began to gain some ground (House, 1993; Patton, 2002; Weiss, 1998). Proponents of this methodological stance highlighted the advantages of qualitative methods for providing a more holistic and in-depth picture of the programme experience, for identifying outcomes not easily captured with standardised quantitative methods (Patton, 2002), and for involving stakeholders and infusing their voices into the evaluation (House, 1993; Patton, 2002; Weiss, 1998). Constructivist (or fourth-generation) evaluation—a form of qualitative evaluation grounded in a relativistic worldview—rejected positivist-imposed evaluation standards and put forward alternatives,

¹ The “utilization crisis” occurred at a complex time in the field’s history that is difficult to capture fully within the scope of this article. Here we only emphasise some of the key debates that characterised the times as these were integral in informing the insights we offer to novel evaluators.

including authenticity criteria, on which the quality of evaluative judgements should be made (Guba & Lincoln, 1989). Nonetheless, these alternative approaches were far from unanimously accepted, and the conflicting viewpoints held by the qualitative-naturalists and the quantitative-positivists generated the long-standing heated debate that came to be known as the “paradigm wars” (Weiss, 1998, p. 14).

Contemporary evaluation practice: Methodological flexibility and diverse evaluation pathways

The debate has now come a long way. There has been a shift towards accepting that phenomena which are difficult to measure and observe directly are nevertheless worth investigating (Trochim & Donnelly, 2008). Debate has largely moved beyond the superiority of process-focused (often qualitative) evaluations versus outcomes-focused (usually quantitative) evaluations, and there is now an emphasis on flexibility. Proponents of mixed-methods approaches (for example, Greene, 2008), explicitly recognise the value that quantitative and qualitative approaches can offer when combined. However, some still argue that while methods may be mixed, methodological paradigms (i.e. positivist and constructivist) cannot be (Kushner, 2002).

Rather than distancing themselves from the evaluand, evaluators are now encouraged to lay their philosophical and values perspective on the table from the outset (Kirkhart, 2015; Patton, 1997, 2002; Trochim & Donnelly, 2008). Stakeholder consultation and at least peripheral involvement is generally seen as good practice (Bamberger et al., 2006; House, 1993; Johnson et al., 2009; Mark, 2003; Patton, 1997; Weiss, 1998). It is now widely accepted that stakeholder involvement also increases evaluation use (Cousins & Chouinard, 2012; Fleischer & Christie, 2009; Johnson et al., 2009).

In addition, the importance of *multicultural validity* (the term used by Kirkhart, 2010) has increasingly been recognised in recent times.

In Aotearoa New Zealand there has been considerable emphasis on working from a bicultural perspective that requires recognition of te ao Māori (the Māori worldview) and tikanga Māori (Māori values). From a Māori perspective, the importance of evaluators connecting spiritually (ā-wairua) with those they are working alongside, and understanding that wairuatanga (spirituality) is inherent to wellness, has been highlighted (Kennedy, Cram, Kirimatao, Pipi, & Baker, 2015). For outsiders to a particular culture, such as being Pākehā in relation to Māori, there is emphasis on the need to increase cultural competence (Torrie, Dalgety, Peace, Roorda, & Bailey, 2015). Accordingly, attention has been drawn to the nature of the “cultural fit” between evaluator and evaluand, and how a disconnect between the culture of the evaluator and evaluand may limit the quality and credibility of an evaluation (Goodwin, Sauni, & Were, 2015).

Conceptualisations of evaluation “use” continue to be an active area for research on evaluation. Arguments that the distinction between “knowledge” and “use” may be somewhat artificial have been advanced. For example, evaluations can reveal new concepts that “usefully” broaden the thinking of practitioners and teach them critical evaluation skills (Patton, 1997), even if this does not produce an immediate change in practice. In addition, Weiss argued (1998) that knowledge production aimed at social-science researchers, policymakers, evaluators, and practitioners involved in similar programmes is legitimately part of an evaluation’s use. This results in “enlightenment”, according to Weiss (1998, p. 24). The concept of “influence” has thus replaced “use” in some evaluation writings to reflect the broad and diffuse means by which evaluations can affect social settings (Henry & Mark, 2003; Kirkhart, 2000; Mark & Henry, 2004).

That stated, debate over the role of the evaluator still exists (Johnson et al., 2009; Kirkhart, 2000; Kushner, 2015; Shadish et

al., 1991; Stufflebeam, 1994; Weiss, 1998). The more participatory approaches (for example, empowerment evaluation, during which stakeholders are supported by an evaluator-facilitator to control the full evaluation process) are still seen by some to lack rigour and credibility as they are thought to reflect the self-interested biases of the stakeholders involved (Stufflebeam, 1994). On the one hand, many theorists maintain that the high degrees of stakeholder involvement seen in some of the evaluation capacity-building approaches should be tempered if the evaluator's role is to render a judgement about the programme's merit and worth (Shadish, 1994; Stufflebeam, 1994; Weiss, 1998). On the other hand, others still strongly oppose the view that evaluation can offer a definitive and generalisable ruling on the worth of any programme, owing to the inherent instability and complexity of the programme context (Kushner, 2015).

In all, then, the simplifications of the “paradigm wars” have been challenged with increasing awareness of the multiple purposes and outcomes of different evaluation approaches. Greater recognition of the diversity in evaluation brings greater awareness of additional tensions that must be carefully navigated. Recognition of these tensions is important for the novice evaluator, but in the words of Cousins, Whitmore and Shulha (2014), this might “leave them feeling ... lost at sea about how to proceed” (p. 150). In such circumstances, a look to a common set of practice principles is a useful next step (Cousins et al., 2014).

The essential features of effective evaluation

In an attempt to move the field forward in a manner that takes account of evaluation's multiple threads while identifying the broadly accepted features of quality evaluation practice, several organisations have outlined basic practice principles and standards. Two of these are the Program Evaluation Standards, originally developed in 1981

by the Joint Committee for Standards on Educational Evaluation (JCSEE) (Yarborough, Shulha, Hopson, & Caruthers, 2011), and the American Evaluation Association's (AEA) Guiding Principles (officially endorsed in 1994; see American Evaluation Association, n.d.). The standards and the principles have had widespread influence and been adapted for the development of frameworks in other regions of the globe.

The features of these frameworks can be seen in first two columns of Table 1. As this table illustrates, they all incorporate the importance of systematic and technically adequate methods; stakeholder consultation and consideration; ethical, respectful, and reflective practice; and detailed, transparent communications. Taken together, these methods are intended to produce both credible and useful social contributions. It is important to note, however, that there is considerable flexibility within each principle and the evaluator must assess how the principles should be applied in any single evaluation.

The local context: Evaluation in Aotearoa New Zealand

Evaluation in Aotearoa New Zealand is strongly influenced by its bicultural foundations. Accordingly, there is a particular emphasis on cultural issues, and while the Aotearoa New Zealand Evaluation Association's (ANZEA) Framework for Evaluator Competencies (ANZEA, 2011) and the Evaluation Standards for Aotearoa New Zealand (Families Commission/Social Policy and Evaluation Research Unit & ANZEA, 2014) exhibit features shared with the JCSEE's Standards and the AEA's principles, there are evident variations from the United States guidelines. This includes an emphasis on cultural responsiveness and commitment to Te Tiriti o Waitangi (The Treaty of Waitangi) including the tripartite principles of partnership, protection, and participation that underpin the bicultural agreement between Māori and Pākehā (see the final two columns in Table 1).

Guidelines for Conducting High-Quality Evaluations			
JCSEE Program Evaluation Standards	AEA Guiding Principles for Evaluators	ANZEA Evaluator Competencies	SUPERU/ANZEA Proposed Evaluation Standards for Aotearoa New Zealand
<p>Utility: Evaluators are credible and qualified; full range of stakeholders attended to; purposes are negotiated and stakeholder needs addressed; explicit articulation of values; information is relevant; processes and products are meaningful; dissemination is timely and appropriate to audiences; concern demonstrated for consequences and influence of evaluation.</p> <p>Feasibility: Effective project management; practical and responsive procedures; diverse cultural and political interests recognised and attended to; effective and efficient use of resources.</p>	<p>Systematic Inquiry: Highest technical standards are adhered to; strengths and opportunities for improvement are explored with client; communications include sufficient details of approach, methods, and limitations to all critique</p>	<p>Contextual analysis and Engagement: Wider context and evaluation situation is identified, understood, articulated, and considered; evaluation team has credibility in the evaluative context and builds respectful relationships with stakeholders; evaluators engage with relevant individuals in wider context when conducting process so interpretations are credible and valid to all affected stakeholders.</p> <p>Systematic Evaluative Inquiry: Evaluator is skilled and knowledgeable in evaluation design, systematic data collection, critical thinking, analysis and synthesis; interpretations are valid and defensible; methods, conclusions and judgements are transparent; findings are reported in credible, useful and actionable way.</p>	<p>Whanaungatanga: Honest, respectful, mana-enhancing relationships; clear, transparent communication of evaluation origin and purpose; accountabilities to all involved are identified; governance between those with authority in evaluation context is negotiated; rights, worldviews, aspirations and desires of evaluation participants are their communities are respected</p> <p>Manakitanga: Cultural and social responsibility; dignity (mana), skills, knowledge, and values of all those involved are upheld, respected and included; ethical protocols to ensure cultural safety, confidentiality, privacy and informed consent are followed; participant contributions are valued and recognised.</p>

JCSEE Program Evaluation Standards	AEA Guiding Principles for Evaluators	ANZEA Evaluator Competencies	SuperU/ANZEA Proposed Evaluation Standards for Aotearoa New Zealand
<p>Propriety: Responsive and inclusive of varied stakeholder groups; formal agreements are negotiated and adhered to; human and legal rights are protected; information is understandable and fair; full disclosure and transparency about process and findings; conflicts of interest are declared; activities are conducted with fiscal responsibility.</p>	<p>Integrity/Honesty: Costs, methods, limitations, scope of results are negotiated in an honest manner; conflicts of interest are disclosed; changes to proposed plan are recorded and reported with implications; interests and values are made explicit; procedures, data, findings are accurately reported and misuse prevented; concerns about misleading information are resolved; financial sources of support and source of request for evaluation are made clear.</p>	<p>Evaluation Project Management and Professional Practice: Evaluator is able to effectively manage the project; to build collaborative and respectful relationships with stakeholders and other team members; appropriate standards and ethics of professional evaluation practice are adhered to.</p>	<p>Methodological Responsiveness & Appropriateness: Methods are appropriate to meet the information needs of stakeholders; process is systematic, rigorous, transparent (including strengths and limitations), feasible and prudent with regards to the available resources and need to produce valuable findings.</p>
<p>Accuracy: Decisions and conclusions are explicitly justified; interpretations are valid; information is dependable and consistent; documentation has sufficient detail of programme and context; technically adequate designs and analytic strategies are employed; evaluative reasoning is explicit; scope of communications is adequate and biases, errors, misconceptions are guarded against</p>	<p>Respect for People: Understanding of broader context is sought; ethical standards are adhered to; benefits are maximised and any potential harms considered and reduced; stakeholder dignity is respected during process and in communications; bidirectional rewards between all those involved are encouraged where possible; diverse stakeholder characteristics and values are considered and respected</p>	<p>Reflective Practice and Professional Development: Evaluator reflects on own areas of expertise, identity and approach to practice; personal growth areas are identified; evaluator engages in professional development opportunities; and contributes to the profession.</p>	<p>Credibility and Competence: Findings and conclusions are sound and have multicultural validity; evaluators have the requisite skills, expertise and experience and exercise independence in decision-making despite conflicts of interest; evaluation contributes to stakeholder decision-making and capacity-building.</p>
<p>Accountability: Evaluations are fully documented; processes and outputs are measured against standards; and external meta-evaluation is encouraged</p>	<p>Responsibilities to General and Public Welfare: Full range of stakeholder perspectives and interests attended to; broad assumptions, implications, and side effects are considered; stakeholders are given access to information and this is actively disseminated in an understandable form; client and stakeholder needs are balanced; interests and welfare of society as a whole are considered</p>		

Table 1. The Essential Features Associated with High Quality Evaluation Practice

At the same time, Nunns et al. (2015) point out that the New Zealand political context influences the type of evidence and methods that are valued. At present, there is resurgence of attention on accountability-focused evaluation for evidence-based policy making (Nunns et al., 2015). Stemming from a strongly positivistic approach, the “evidence-based movement” has been acknowledged to work in opposition to ecological approaches that embrace complexity, diversity, respectful relationships, and localised solutions that promote community empowerment (Trickett, 2015). And so, in the political context of Aotearoa New Zealand, evaluators find themselves faced with demands for “evidence” that must be balanced with an increasingly sophisticated understanding of what social settings, bicultural values, and influential evaluations actually entail. This is difficult for any evaluator to negotiate; it may leave novice evaluators unable to even take the first step.

Moving from evaluation theory to thoughtful evaluation practice

What this history suggests is that while particular issues come and go, there are two broad dimensions that underpin many of the theoretical debates and practical tensions evaluators experience. On the one hand is the importance of *methodological credibility*; on the other is the importance of *stakeholder empowerment*. Features within the methodological credibility dimension align with concepts of research integrity that one would find in any good research-methods text (for example, Trochim & Donnelly, 2008) and in manuals designed to aid rigorous evaluations (for example, Bamberger, et al., 2006). They direct attention to the construction of a defensible and valid evaluative judgement, and align strongly with the formal guidelines outlined earlier that include *systematic inquiry*, *accuracy*, and *methodological appropriateness* (see Table 1).

According to the community psychology perspective that we draw on, stakeholder empowerment, in contrast, focuses on an increase in the stakeholders' knowledge and decision-making skills, as well as their ability to participate in and learn from the evaluation and shape the social context in response to what is learnt (see Fetterman & Wandersman, 2005; Zimmerman, 2001). The stakeholder empowerment dimension aligns strongly with the inclusion of *utility, feasibility, propriety, respect for people, responsibilities to general and public welfare, contextual analysis and engagement, project management and professional practice, and whanaungatanga and manaakitanga* (see Table 1) in the formal guidelines (ANZEA, 2011; Families Commission/Social Policy and Evaluation Research Unit & ANZEA, 2014; Yarborough, et al., 2011).

In the short term, these two dimensions often pull against each other. To illustrate: an outcome evaluation conducted by an external evaluator may be viewed as high in methodological credibility but low in stakeholder empowerment if the stakeholders' concerns and interests are not considered in the evaluation design. In contrast, an evaluation that allows stakeholders (especially those who implement the programme) full control of the process may be empowering for those in control but may have limited credibility beyond the stakeholders themselves. This is especially true if the stakeholders are not well-versed in formal research methods. However, in the longer term—and as will be illustrated by the case study described next—for a method to be credible and for stakeholders to be empowered, these two dimensions should both be held in mind and be allowed to enhance each other.

To negotiate these tensions, we suggest a novice evaluator take a *thoughtful* approach, which involves four key aspects. First, the thoughtful evaluator needs to decide what values she or he wants to advance. By articulating the “values dimensions of our craft”

(Greene, 2011, p. 8) evaluators have a relatively stable point from which to make further decisions; including whether to undertake a particular evaluation. Secondly, the thoughtful evaluator needs some understanding of the different research designs that can contribute to methodologically credible evaluations. It is not expected or possible for a novice (or an experienced!) evaluator to be familiar with the complete range of approaches, but we suggest it is important to have some awareness of what is possible and the contexts in which different methods may be appropriate. If it is clear that a particular evaluation requires an approach that is, for whatever reason, outside the capabilities of the evaluator, then this needs to be accepted, rather than an inappropriate approach applied (for instance, if there is a substantial discrepancy in the cultural fit between evaluator and evaluand, as suggested by Goodwin et al., 2015). Thirdly, the thoughtful novice should endeavour to understand the social context of the evaluation and, in particular, the issues around stakeholder involvement that may facilitate or detract from empowerment. Finally, the evaluator will develop professionally and be more effective if she or he engages in ongoing self-reflection as the evaluation unfolds in view of his or her values, knowledge, and experience. This involves closely examining one's assumptions and a commitment to continually learning from experience, as is the case with other reflective practice (see Schön, 1983). While there are various strategies for reflection, regular consultation with a mentor is strongly recommended. This is a common requirement of professional practice in other professions, and while not often discussed in relation to evaluation practice, we feel it should be. In Aotearoa New Zealand adequate self-reflection also includes listening to different cultural voices, engaging with different cultural perspectives and embarking on one's own journey of increased cultural humility.

We now illustrate this approach in practice. The first author began as a novice evaluator under the mentorship of the second author (in a doctoral student/supervisor relationship). The authors developed the thoughtful evaluation approach and then used it to work alongside a youth development organisation (called the Foundation for Youth Development (FYD) until 2016, now called the Graeme Dingle Foundation), first in the evaluation of their flagship programme (Project K), and then in a series of further evaluations and discussions of programme development.

Reflecting on one's vision in practice:

Thoughtful evaluation in action

FYD provides governance, research, and training for licensed community partners to deliver five youth development programmes across many sites throughout New Zealand. The Project K programme targets Year 10 students with low self-efficacy and engages them in outdoor adventure, community, and mentoring activities. In 2004, a randomised controlled trial (RCT) outcome evaluation of the programme was initiated with the support of the second author and the Ministry of Social Development. The first author partnered with FYD in 2007 to work on the RCT and conduct a broader evaluation of Project K. This evaluation partnership was the basis of her doctoral thesis conducted under the supervision of the second author. We now explain how the first author, as a novice evaluator, used the thoughtful-evaluation approach to evaluate Project K and enhance the learning of all parties.

First, the underlying values of FYD were highly compatible with those of the first author (thoughtful evaluation feature one), as she sees considerable value in promoting the holistic development of young people through programmes and, in particular, FYD's intensive, well-structured, and people-oriented approach and commitment

to organisational learning through ongoing evaluation. Second, as a PhD student, the first author undertook a literature review and training that increased her knowledge of methodological options (thoughtful evaluation feature two). Importantly, the programme designers and FYD's research and evaluation team were fully supportive of a methodologically credible evaluation and had provided considerable resources to enable this through the RCT. While an RCT may, and in this case did, have several limitations, its instigation nevertheless reflected the organisation's openness and interest in this core facet of evaluation. Nevertheless, the budget was tight and the structure of the evaluation meant that programme deliverers (embedded in the community partner organisations) needed to gather much of the evaluation data. It was also apparent from evaluation training sessions that the programme deliverers struggled with their role in the evaluation at times.

The first author therefore initiated discussions with programme deliverers (consistent with thoughtful evaluation feature three), and found they were not always clear if and why the evaluation was needed. Consequently, the programme deliverers were sometimes reluctant to make data collection a priority. It became apparent that stakeholder empowerment was critical for the success of the RCT, but also had been somewhat compromised by the RCT. In an effort to redress this balance, the first author put increased effort into building strong, respectful, and reciprocal relationships with the programme deliverers, including setting-up sessions to discuss and vote for evaluation priorities in relation to the RCT data and to other potential evaluation questions. These sessions also solicited feedback on the preliminary findings of the RCT.

However, relationship building takes considerable time, which is often scarce. In this case, additional time was involved in reaching programme sites spread throughout the country. The first author

found it difficult to schedule additional meetings or workshops that would involve programme deliverers more fully in the evaluation process. She also discovered that just providing opportunities was not enough to secure stakeholder engagement with the evaluation, and that evaluators have no right to demand this engagement. In addition, she found it necessary to forego some of her ambitions for involving participants and their families more directly in the evaluation. This was because participants and families would need to be recruited via the programme deliverers who did not have time to facilitate this process.

Ironically, conducting an empowering evaluation requires a diverse representation of stakeholders to be involved, and to want to be involved. Active reflection in supervision discussions (thoughtful evaluation feature four) emphasised the gap between the first author's hope for a robust RCT evaluation and enthusiastic participation from programme deliverers who would further tailor the evaluation questions and her experiences in practice. This enabled us to surface the realisation that, as evaluators, we can offer opportunities for stakeholders to voice and potentially seize power, but we cannot force empowerment, as this goes against its very notion. We called this the *empowerment paradox*. Importantly, our experience resonates with critiques of empowerment articulated by others. For instance, Luttrell, Quiroz, Scrutton, and Bird (2009) highlight the multidimensional nature of power and questions outsiders who attempt to “empower” others, given that this may inadvertently involve manipulating the current beliefs and values of those perceived to be disempowered in a manner that may in fact be disempowering. Both Labonte (1994) and Luttrell et al. (2009) illustrate that, without addressing the systemic causes of disempowerment, attempts to increase individual empowerment via capacity-building may be futile. In our case, the emphasis on gathering RCT

data—driven in many ways by the political environment discussed earlier that values “evidence-based outcomes” over and above institutional learning—had potentially detracted from efficient data gathering because it had left the key stakeholders behind. However, simply inviting stakeholders to become involved in the evaluation was not empowering because the broader systemic issues to do with their lack of time to be more involved were not addressed. It was evident that a change in organisational culture would be needed at a much broader level to be genuinely empowering. All of this points to the importance of critical self-reflection when approaching an evaluation with an “empowerment” agenda.

Eventually, through reflection, more discussions with FYD, and further consultation of the evaluation literature, the first author was able to reframe her perspective of what would actually allow programme deliverers to have a meaningful voice within the evaluation process, without requesting they participate beyond their capacity or interest. One realisation was that the programme deliverers needed to be *less* not *more* involved in gathering data for the RCT evaluation, but have increased opportunity to feed into programme analysis and development. To do this, she worked with FYD to develop a programme theory of change based on staff interviews, qualitative comments from youth participants, and programme document analysis (Deane & Harré, 2014).

Keeping abreast of developments in the local evaluation sector (for example, the introduction of proposed Evaluator Competencies in Aotearoa) allowed the first author (an immigrant to New Zealand) to recognise an additional shortcoming of her approach with regards to Māori cultural responsiveness. Though this realisation occurred late in the evaluation process, this encouraged her to seek support from a Māori cultural advisor, who agreed to review her interpretations with a critical cultural lens. This feedback enhanced the cultural sensitivity

of the interpretations, as did her later engagement in tikanga and te reo classes.

Although there were challenges and shortcomings in this first attempt to put a thoughtful approach into practice, there were also many successes, and a deep ongoing relationship between our university and FYD is still unfolding. We were able to conduct an analysis of the RCT data enhanced by a broad range of questions submitted by the programme deliverers in the process described earlier. This moved understanding of programme effects beyond a simple pre–post account of average changes on blanket outcomes for Project K and control participants to a more nuanced investigation of effects by region, gender, ethnicity, and school decile. Consequently, this provided a deeper understanding of who the programme works best for (at least with regards to the variables measured; see Deane, Harré, Moore, & Courtney, 2016). The theory of change analysis also enabled us to identify new evaluation questions that were not considered by the RCT, and to explore additional methodological approaches such as interviews and observations (Deane & Harré, 2014).

The respectful relationships developed with FYD and the programme deliverers also enabled other university researchers and students to work on evaluation projects with Project K and the organisation's other programmes. These incorporated the learnings from the initial evaluation. For instance, a recently completed 2-year evaluation of the impact of FYD's Stars programme on peer mentors involved: initial consultation with Community Partners and mentors on the evaluation design and measures that would work best given their interests and pragmatic constraints; regular dissemination to partner schools, staff, and participants; workshop opportunities for the youth participants to assist us with the interpretation of the results and increase their knowledge of evaluation research (Deane, Moore, Gillham & Brown, 2015); and a programme development

partnership project informed by the findings and involving the first author's undergraduate students.

The first author's reflections on the experience with the Project K RCT evaluation also informed the development of a staged exploratory evaluation workshop process informed by a blend of evaluation models—utilisation-focused (Patton, 1997), empowerment (Fetterman & Wandersman, 2005), and theory-driven (Donaldson, 2007)—as well as systems thinking (Williams & Hummelbrunner, 2010), and evaluability assessment (Wholey, 2010). The purpose of the process is to involve programme deliverers and other stakeholders intimately in the evaluation design process in short but intense sessions at times that work with their busy schedules, and cultivate deeper levels of evaluative thinking. This informs the design of subsequent evaluation work (Deane & Bullen, 2015). Transparent discussions about intentions, expectations, and written collaborative agreements help to mitigate some of the challenges associated with the empowerment paradox outlined above. Anonymous feedback solicited from workshop participants suggest that it is enlightening, affirming, and useful; these themes resonate strongly with notions of individual empowerment.

To sum up: the first author engaged with FYD due, in the first instance, to a sensed compatibility between her values and theirs, and this sense has never wavered. She stayed committed to producing methodologically credible research as well as contributing to an empowering context for programme deliverers. Although there were many tensions, her process of reflection with the help of her mentor, the second author, enabled meaningful evaluations and institutional learning to occur. The thoughtful approach is now “second nature” to her as she continues to develop as an evaluator.

Conclusion

Programme evaluation is a values-driven process. First, every social programme is underpinned by the values of its stakeholders and the socio-political context in which it is embedded. Secondly, the evaluator brings another layer of expectations and attempts to judge the worthiness of the endeavour. The *thoughtful approach* to evaluator development is based on a brief analysis of historical evaluation debates, with specific attention to the context of Aotearoa New Zealand. We offer the insights generated from our experience to novice evaluators interested in advancing the integrity of their own practice.

Acknowledgements

Portions of this work were presented in thesis form in fulfilment of the requirements for the PhD of Kelsey L. Deane from the University of Auckland. The research was supported by a New Zealand International Doctoral Research Scholarship and a University of Auckland International Doctoral Research Scholarship. We would like to thank the editor of *Evaluation Matters—He Take Tō Te Aromatawai* and the two blind reviewers for their helpful feedback on earlier versions of this article. We would also like to thank the Graeme Dingle Foundation (previously FYD) for supporting our desire to publish this article using our partnership experience as a case example.

References

- American Evaluation Association (n.d.). *American Evaluation Association guiding principles for evaluators*. Retrieved from <http://www.eval.org/p/cm/ld/fid=51>
- Aotearoa New Zealand Evaluation Association. (2011). *Evaluator competencies*. Retrieved from http://anzea.org.nz/images/documents/110801_anzea_evaluator_competencies_final.pdf
- Bamberger, M., Rugh, J., & Mabry, L. (2006). *RealWorld evaluation: Working*

- under budget, time, data, and political constraints.* Thousand Oaks, CA: Sage.
- Chen, H.-T. (1990). *Theory-driven evaluations.* Newbury Park, CA: Sage Publications.
- Cousins, J. B. & Chouinard, J. A. (2012). *Participatory evaluation up close: An integration of research-based knowledge.* Charlotte, NC: Information Age Publishing.
- Cousins, J. B., Whitmore, E. & Shulha, L. (2014). Let there be light: Response to Fetterman et al., 2014. *American Journal of Evaluation*, 35(1), 149–153. <http://dx.doi.org/10.1177/1098214013510200>
- Deane, K. L., & Bullen, P. (2015). *Choose your own evaluation adventure: A staged workshop for designing a contingency-based program evaluation.* Workshop presented at the 15th Society for Research and Action Biennial Conference, Lowell, MA.
- Deane, K. L., & Harré, N. (2014). Program theory-driven evaluation science in a youth development context. *Evaluation and Program Planning*, 45, 61–70. <http://dx.doi.org/10.1016/j.evalprogplan.2014.03.009>
- Deane, K. L., Harré, N., Moore, J., & Courtney, M. G. (2016). The impact of the Project K youth development program on self-efficacy: A randomized controlled trial. *Journal of Youth and Adolescence*. <http://dx.doi.org/10.1007/s10964-016-0463-9>
- Deane, K. L., Moore, J., Gillham, B., & Brown, K. (2015). *The impact of Stars on Peer Mentors: A report for the New Zealand Lottery Grants Board.* Retrieved from: <http://www.communityresearch.org.nz/research/the-impact-of-stars-on-peer-mentors-2/>
- Donaldson, S. I. (2007). *Program theory-driven evaluation science: Strategies and applications.* New York, NY: Lawrence Erlbaum Associates.
- Families Commission/Social Policy and Evaluation Research Unit & Aotearoa New Zealand Evaluation Association (2014). *Proposed evaluation standards for Aotearoa New Zealand.* Wellington, NZ: Authors. Retrieved from <http://www.anzea.org.nz/wp-content/uploads/2014/08/FMC562-Evaluation-Standards-A.pdf>

- Fetterman, D. & Wandersman, A. (2005). *Empowerment evaluation: Principles in practice*. New York, NY: Guildford Press.
- Fleischer, D. N. & Christie, C. A. (2009). Evaluation use: Results from a survey of U.S. American Evaluation Association members. *American Journal of Evaluation*, 30(2), 158–175.
<http://dx.doi.org/10.1177/1098214008331009>
- Goodwin, D., Sauni, P., & Were, L. (2015). Cultural fit: An important criterion for effective interventions and evaluation work. *Evaluation Matters—He Take Tō Te Aromatawai*, 1, 25–46.
<http://dx.doi.org/10.18296/em.0003>
- Greene, J. C. (2008). Is mixed methods social inquiry a distinctive methodology? *Journal of Mixed Methods Research*, 2(1), 7–22.
<http://dx.doi.org/10.1177/1558689807309969>
- Greene, J. C. (2011). *Evaluation and the public good: A mixed methods perspective*. Paper presented at the Aotearoa New Zealand Evaluation Association Conference, Wellington, New Zealand.
- Guba, E. G. & Lincoln, Y. S. (1989). *Fourth generation evaluation*. Newbury Park, CA: Sage Publications.
- Henry, G. T. & Mark, M. M. (2003). Beyond use: Understanding evaluation's influence on attitudes and actions. *American Journal of Evaluation*, 24(3), 293–314.
<http://dx.doi.org/10.1177/109821400302400302>
- House, E. R. (1993). *Professional evaluation: Social impact and political consequences*. Newbury Park, CA: Sage.
- Johnson, K., Greenesid, L. O., Toal, S. A., King, J. A., Lawrenz, F. & Volkov, B. (2009). Research on evaluation use: A review of the empirical literature from 1986 to 2005. *American Journal of Evaluation*, 30(3), 377–410.
<http://dx.doi.org/10.1177/1098214009341660>
- Kennedy, V., Cram, F., Kirimatao, P., Pipi, K., & Baker, M. (2015). Wairua and cultural values in evaluation. *Evaluation Matters—He Take Tō Te Aromatawai*, 1, 83–111. <http://dx.doi.org/10.18296/em.0005>

- Kirkhart, K. E. (2000). Reconceptualizing evaluation use: An integrated theory of influence. *New Directions for Evaluation*, 88, 5–23.
<http://dx.doi.org/10.1002/ev.1188>
- Kirkhart, K. E. (2010). Eyes on the prize: Multicultural validity and evaluation theory. *American Journal of Evaluation*, 31(3), 400–413.
<http://dx.doi.org/10.1177/1098214010373645>
- Kirkhart, K. E. (2015). Unpacking the evaluator's toolbox: Observations on evaluation, privilege, equity and justice. *Evaluation Matters—He Take Tō Te Aromatawai*, 1, 7–24. <http://dx.doi.org/10.18296/em.0002>
- Kushner, S. (2002). I'll take mine neat: Multiple methods but a single methodology. *Evaluation*, 8(2), 249–258.
<http://dx.doi.org/10.1177/1358902002008002521>
- Kushner, S. (2015). Chasing the McGuffin: Towards the continuing reform of program evaluation. *Evaluation*, 21(2), 204–215.
<http://dx.doi.org/10.1177/1356389015577467>
- Labonte, R. (1994). Health promotion and empowerment: Reflections on professional practice. *Health Education Quarterly*, 21(2), 253–268.
<http://dx.doi.org/10.1177/109019819402100209>
- Lipsey, M. W., & Cordray, D. S. (2000). Evaluation methods for social intervention. *Annual Review of Psychology*, 51(1), 345–375.
<http://dx.doi.org/10.1146/annurev.psych.51.1.345>
- Luttrell, C., Quiroz, S., Scrutton, C., & Bird, K. (2009). *Understanding and operationalising empowerment* (Working paper 308). London: Overseas Development Institute.
- Mark, M. M. (2003). Toward an integrative view of the theory and practice of program and policy evaluation. In S. I. Donaldson & M. Scriven (Eds.), *Evaluating social programs and problems: Visions for the new millennium*. (pp. 183–204). Mahwah, NJ: Lawrence Erlbaum Associates.
- Mark, M. M., & Henry, G., (2004). The mechanisms and outcomes of evaluation influence. *Evaluation*, 10(1), 35–57.
<http://dx.doi.org/10.1177/1356389004042326>

- Nunns, H., Peace, R., & Witten, K. (2015). Evaluative reasoning in public-sector evaluation in Aotearoa New Zealand: How are we doing? *Evaluation Matters—He Take Tō Te Aromatawai, 1*, 137–163.
<http://dx.doi.org/10.18296/em.0007>
- Patton, M. Q. (1997). *Utilization-focused evaluation: The new century text* (3rd ed.). Thousand Oaks, CA: Sage.
- Patton, M. Q. (2002). *Qualitative research & evaluation methods* (3rd ed.). Thousand Oaks, CA: Sage.
- Rossi, P. H., & Freeman, H. E. (1989). *Evaluation. A systematic approach* (4th ed.). Newbury Park, CA: Sage.
- Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M., Gagne & M. Scriven (Eds.), *Perspectives of curriculum evaluation* (pp. 31–83). Chicago, IL: Rand McNally.
- Schön, D. A. (2003). *The reflective practitioner: How professionals think in action*. Aldershot, England: Ashgate Publishing.
- Shadish, W. R. (1994). Need-based evaluation theory: What do you need to know to do good evaluation? *Evaluation Practice, 15*(3), 347–358.
[http://dx.doi.org/10.1016/0886-1633\(94\)90029-9](http://dx.doi.org/10.1016/0886-1633(94)90029-9)
- Shadish, W. R., Cook, T. D., & Leviton, L. C. (1991). *Foundations of program evaluation: Theories of practice*. Newbury Park, CA: Sage.
- Stufflebeam, D. L. (1994). Empowerment evaluation, objectivist evaluation, and evaluation standards: Where the future of evaluation should not go and where it needs to go. *Evaluation Practice, 15*(3), 321–338.
[http://dx.doi.org/10.1016/0886-1633\(94\)90027-2](http://dx.doi.org/10.1016/0886-1633(94)90027-2)
- Torrie, R., Dalgety, M., Peace, R., Roorda, M. & Bailey, R. (2015). Finding our way: Cultural competence and Pākehā evaluators. *Evaluation Matters—He Take Tō Te Aromatawai, 1*, 47–81.
<http://dx.doi.org/10.18296/em.0004>
- Trickett, E. J. (2015). Seymour Sarason remembered: “Plus ça change...”, “Psychology Misdirected”, and “Community Psychology and the Anarchist Insight”. *American Journal of Community Psychology, 56*(3),

- 197-204. <http://dx.doi.org/10.1007/s10464-015-9744-9>
- Trochim, W. M. K., & Donnelly, J. P. (2008). *The research methods knowledge base* (3rd ed.). Mason, OH: Cengage Learning.
- Weiss, C. H. (1998). *Evaluation: Methods for studying programs and policies* (2nd ed.). Upper Saddle River, NJ: Prentice-Hall.
- Wholey, J. (2010). Exploratory evaluation. In J. S. Wholey, H. P. Hatry, & K. E. Newcomer (Eds.), *Handbook of practical evaluation* (3rd ed.), (pp. 81–99). San Francisco, CA Jossey-Bass.
- Williams, B. & Hummelbrunner, R. (2010). *Systems concepts in action: A practitioner's toolkit*. Stanford, CA: Stanford University Press.
- Yarbrough, D. B., Shulha, L. M., Hopson, R. K., & Caruthers, F. A. (2011). *The program evaluation standards: A guide for evaluators and evaluation users* (3rd ed.). Thousand Oaks, CA: Sage.
- Zimmerman, M. A. (2001). Empowerment theory: Psychological, organizational, and community level analysis. In J. Rappaport & E. Seidman (Eds.), *Handbook of Community Psychology* (pp. 43–63). New York, NY: Kluwer Academic.

The authors

Kelsey Deane, PhD, is a lecturer in the School of Counselling, Human Services and Social Work, Faculty of Education and Social Work at The University of Auckland. Her major research interests include positive youth development, youth mentoring and program evaluation.

Email: k.deane@auckland.ac.nz

Niki Harré, PhD, is an associate professor in the School of Psychology in the Faculty of Science at The University of Auckland. She teaches and researches in community psychology and the psychology of sustainability.

Email: n.harre@auckland.ac.nz