The Canadian Journal of Program Evaluation | La Revue canadienne d'évaluation de programme

34.2 Fall 2019 / automne 2019

Editor's Remarks / Un mot de la rédactrice

vii

Isabelle Bourgeois

ARTICLES

Revisiting Contribution Analysis

171

John Mayne

Self-Evaluation Tool for Action in Partnership: Translation and Cultural Adaptation of the Original Quebec French Tool to Canadian English

192

Angèle Bilodeau and Gillian Kranias

Comparison of Canadian and American Graduate Evaluation Education Programs

207

M. Theresa Hunter and James C. McDavid

Évaluation de la mise en œuvre du programme d'intervention *In vivo* chez les jeunes de 11–12 ans d'une école primaire québécoise présentant un problème de comportements intériorisés : une étude exploratoire

235

Catherine Fréchette-Simard, Jonathan Bluteau et Isabelle Plante

Canadian
Evaluation
Society

Société
canadienne
d'évaluation

RESEARCH AND PRACTICE NOTES / NOTES SUR LA RECHERCHE ET LES MÉTHODES

Dramatizing Learning, Performing Ourselves: Stories of
Theatre-Based Evaluation in Vancouver's Downtown Eastside 255
*Christopher Cook, Carolyn Camman, Andrew Leyland,
Suzie O'Shea, and Angela Towle*

Applying the Collaborative Approaches to Evaluation (CAE)
Principles in an Educational Evaluation: Reflections from the Field 272
Jeremy Acree

THEME SEGMENT: DEMOCRACY AND EVALUATION / SEGMENT THÉMATIQUE : LA DÉMOCRATIE ET L'ÉVALUATION

Démocratisation et efficience sont-elles antinomiques au niveau
local? Le cas des réformes managériales des administrations 282
publiques locales en Wallonie
Catherine Fallon

"Deliverology" and Evaluation: A Tale of Two Worlds 303
Lisa Birch and Steve Jacob

THEME SEGMENT: 2018 CES KEYNOTE PANEL / SEGMENT THÉMATIQUE : CONFÉRENCES INVITÉES LORS DU CONGRÈS DE 2018

Introduction to Articles Prepared by CES Calgary 2018 Keynote
Panel Members on *Reconciliation and Culturally Responsive*
Evaluation—Rhetoric or Reality? 329
Debbie DeLancey

Creating New Stories: The Role of Evaluation in Truth
and Reconciliation 331
Larry Bremner

Nation-to-Nation Evaluation: Governance, Tribal Sovereignty,
and Systems Thinking through Culturally Responsive 343
Indigenous Evaluations
Nicole Bowman (Waapalaneexkweew)

White Privilege and the Decolonization Work Needed in Evaluation
to Support Indigenous Sovereignty and Self-Determination 357
Kate McKegg

Increasing Cultural Competence in Support of
Indigenous-Led Evaluation: A Necessary Step 368
toward Indigenous-Led Evaluation
Nan Wehipeihana

BOOK REVIEWS / COMPTES RENDUS DE LIVRES

Evaluation Use and Decision Making in Society: A Tribute to Marvin C. Alkin, edited by Anne Vo and Christina A. Christie 385
Reviewed by Margaret Schultz Patel / compte-rendu par Margaret Schultz Patel

Monitoring and Evaluation Training: A Systematic Approach, by Scott G. Chaplowe and J. Bradley Cousins 388
Reviewed by Michele Tarsilla / compte-rendu par Michele Tarsilla

Editorial Team / Équipe éditoriale

**Editor-in-chief /
Rédactrice en chef**

Isabelle Bourgeois, Ph.D.
Professeure agrégée/Associate Professor
Université d'Ottawa/University of Ottawa
(613) 562-5800 ext. 4291
<isabelle.bourgeois@uottawa.ca>

**Associate Editor /
Rédactrice associée**

Jill Anne Chouinard, Ph.D.
Assistant Professor, University of Victoria
<jill.chouinard@uncg.edu>

**Book Review Editor /
Rédactrice, Comptes
rendus de livres**

Jane Whynot
University of Ottawa
<jwhyn021@uottawa.ca>

**Editorial Assistant /
Adjointe à la rédaction**

Stéphanie Maltais, Ph.D.
<smalt006@uottawa.ca>

EDITORIAL BOARD / COMITÉ DE RÉDACTION

Courtney Amo	Atlantic Canada Opportunities Agency
Tim Aubry	School of Psychology, University of Ottawa
Ayesha Boyce	University of North Carolina at Greensboro
Bernadette Campbell	Carleton University
Brad Cousins	Faculty of Education, University of Ottawa
Sarah Earl	YMCA GTA
Paul Favaro	Peel Board of Education
Marie Gervais	Université Laval
Steve Jacob	Université Laval
Marlène Laeubli	Switzerland
Chris Lovato	University of British Columbia
John Mayne	Consultant, Ottawa
James McDavid	School of Public Administration, University of Victoria
Céline Mercier	Centre de réadaptation Lisette Dupras, Lachine
Anita Myers	Department of Health Studies, University of Waterloo
Michael Obrecht	Consultant, Ottawa
John Owen	Centre for Program Evaluation, University of Melbourne
Burt Perrin	Consultant, France
Cheryl Poth	University of Alberta
Hallie Preskill	FSG Social Impact Consultants, San Francisco
Linda Rey	One Drop
Lucie Richard	Faculté des sciences infirmières, Université de Montréal
Valéry Ridde	Université de Montréal
Ray C. Rist	The World Bank, Washington, D.C.
Daniela Schröter	Western Michigan University
Robert Schwartz	University of Toronto
Mark Seasons	School of Planning, University of Waterloo
Souraya Sidani	Ryerson University
Nick L. Smith	School of Education, Syracuse University
Sanjeev Sridharan	Health Policy Management & Evaluation, University of Toronto

PRINTED AND BOUND IN CANADA

PUBLICATIONS MAIL AGREEMENT NO 40600510

RETURN UNDELIVERABLE CANADIAN ADDRESSES TO

UNIVERSITY OF TORONTO PRESS, JOURNALS DIVISION

5201 DUFFERIN ST., TORONTO, ONTARIO, M3H 5T8

EMAIL: journals@utpress.utoronto.ca

Editor's Remarks

I am pleased to present this issue of *CJPE*, which contains two theme segments as well as several practice notes and book reviews. The first theme segment, which includes an article by Catherine Fallon in French and an article by Lisa Birch and Steve Jacob in English, was guest edited by Jacob and focuses on democracy and evaluation. The second theme segment features a series of keynote addresses on reconciliation and culturally responsive evaluation from the 2018 CES National Conference and was guest edited by Nan Wehipeihana. Our keynote speakers have generously put their addresses into writing in order to reach all CES members and other evaluators around the world. I would like to take this opportunity to thank Nicole Bowman-Farrell, Larry Bremner, Debbie Delancey, Kate McKegg, and Nan Wehipeihana for their contributions. In addition to these two segments, this issue also features an important update on contribution analysis by John Mayne, an innovative approach to partnership evaluation by Angèle Bilodeau and Gillian Kranias, an inventory of Canadian evaluation education programs by Theresa Hunter and James McDavid, and an evaluation exemplar by Catherine Fréchette-Simard, Jonathan Bluteau and Isabelle Plante. This issue also includes a research and practice note by Jeremy Acree on the collaborative approaches to evaluation principles, and one by Christopher Cook and collaborators on theatre-based evaluation. Finally, we include two book reviews. I thank our authors for their continued submissions and contributions and hope that this issue provides thought-provoking material for all of our readers.

Isabelle Bourgeois
Editor-in-Chief

Un mot de la rédactrice

Je suis heureuse de vous présenter ce numéro de la RCEP, qui a deux segments thématiques, plusieurs notes sur la pratique ainsi que des comptes-rendu de livres à vous offrir. Le premier segment thématique, qui comprend un article de Catherine Fallon en français et un article de Lisa Birch et Steve Jacob en anglais, pour lequel M. Jacob a été le rédacteur invité, porte sur la démocratie et l'évaluation. Le deuxième segment thématique est composé de textes issus des conférences invitées sur la réconciliation et l'évaluation axée sur la culture présentées lors du Congrès national de la SCE en 2018; Nan Wehipeihana en est la rédactrice invitée. Nos conférencières et conférenciers ont généreusement transcrit et mis à jour leurs exposés pour en faire part à tous les membres de la SCE, et à d'autres évaluateurs et évaluatrices de partout dans le monde. J'aimerais profiter de l'occasion pour remercier Nicole Bowman-Farrell, Larry Bremner, Debbie Delancey, Kate McKegg et Nan Wehipeihana pour leurs contributions. En plus de ces deux segments, le présent numéro propose une importante mise à jour sur l'analyse de contribution préparée par John Mayne; un outil pour évaluer les partenariats, présenté par Angèle Bilodeau et Gillian Kranias; un inventaire des programmes canadiens de formation en évaluation réalisé par Theresa Hunter et James McDavid; et un modèle d'évaluation développé par Catherine Fréchette-Simard, Jonathan Bluteau et Isabelle Plante. Ce numéro comprend également une note sur la pratique et la recherche de Jeremy Acree au sujet des principes visant à guider les approches collaboratives d'évaluation, ainsi qu'une note de Christopher Cook et ses collaborateurs sur l'évaluation fondée sur le théâtre. Finalement, vous trouverez aussi deux comptes-rendu de livres. Je remercie les auteur.e.s qui continuent à nous envoyer leurs soumissions et leurs contributions et j'espère que le présent numéro saura générer des réflexions enrichissantes pour tous nos lecteurs et toutes nos lectrices.

*Isabelle Bourgeois
Rédactrice en chef*

Revisiting Contribution Analysis

John Mayne

Abstract: *The basic ideas behind contribution analysis were set out in 2001. Since then, interest in the approach has grown and contribution analysis has been operationalized in different ways. In addition, several reviews of the approach have been published and raise a few concerns. In this article, I clarify several of the key concepts behind contribution analysis, including contributory causes and contribution claims. I discuss the need for reasonably robust theories of change and the use of nested theories of change to unpack complex settings. On contribution claims, I argue the need for causal narratives to arrive at credible claims, the limited role that external causal factors play in arriving at contribution claims, the use of robust theories of change to avoid bias, and the fact that opinions of stakeholders on the contribution made are not central in arriving at contribution claims.*

Keywords: *causal factors, causal narratives, contribution analysis, contribution claims, contributory causes, theories of change*

Résumé : *Les principes fondamentaux de l'analyse de contribution ont été établis en 2001. Depuis lors, l'intérêt porté à cette approche a crû et l'analyse de contribution a été opérationnalisée de différentes façons. De plus, plusieurs examens de cette approche ont été publiés et ont soulevé quelques inquiétudes. Dans cet article, je clarifie plusieurs concepts de l'analyse de contribution, incluant les causes contributives et les énoncés de contribution. Je discute de la nécessité de faire appel à des théories du changement raisonnablement robustes et d'utiliser des théories complémentaires du changement pour comprendre des contextes complexes. Au chapitre des énoncés de contribution, je soutiens qu'il est nécessaire d'élaborer des narratifs causaux pour arriver à des attributions crédibles; que les facteurs causaux externes jouent un rôle limité dans l'atteinte des énoncés de contribution; que l'utilisation de théories du changement robustes permet d'éviter les biais; et que les opinions des intervenant.e.s sur la contribution ne devraient pas jouer un rôle central dans l'établissement des énoncés de contribution.*

Mots clé : *facteurs causaux, narratifs causaux, analyse de contribution, attributions de contribution, causes contributives, théories du changement*

Increasingly, interventions that evaluators are asked to assess are quite complicated and complex. They may involve a number of major components, different levels of government and/or numerous partners, and have a long timeframe, perhaps with

Corresponding author: John Mayne, john.mayne@rogers.com

emerging outcomes (Byrne, 2013; Gerrits & Verweij, 2015; Schmitt & Beach, 2015). Nevertheless, funders of such interventions still want to know if their funding has made a difference—if the interventions have improved the lives of people—and in what manner. While a range of evaluation approaches might address these questions, theory-based methods are often used, including contribution analysis (Befani & Mayne, 2014; Befani & Stedman-Bryce, 2016; Mayne, 2001, 2011, 2012; Paz-Ybarnegaray & Douthwaite, 2016; Punton & Welle, 2015; Stern et al., 2012; Wilson-Grau & Britt, 2012).

Contribution analysis (CA) has continued to evolve since its introduction in 2001 (Budhwani & McDavid, 2017; Dybdal, Nielsen, & Lemire, 2010). It was first presented in the setting of using monitoring data to say something about causal issues related to an intervention. Since then, most thinking about and application of CA has been as an evaluation approach to get at causal issues and understanding about how change is brought about. At the same time, my concepts and ideas about theories of change—the basic tool used for CA—have evolved considerably (Mayne, 2015, 2017, 2018). In this article, I would like to set out my current thinking on several key issues and some misunderstandings around CA:

- how causality is understood and addressed in CA,
- useful theories of change for CA in complex settings,
- inferring causality for contribution claims, and
- generalizing CA findings on contribution claims.

Those using or reviewing contribution analysis have raised several concerns about its application (Budhwani & McDavid, 2017; Delahais & Toulemonde, 2012, 2017; Dybdal et al., 2010; Lemire, Nielsen, & Dybdal, 2012; Schmitt & Beach, 2015). I will address these concerns and issues as the article unfolds. The article aims to correct a number of misinterpretations around CA. It builds on several previous publications and assumes some working knowledge of CA.

First, here is a review of the terms being used:

- **Impact pathways** describe causal pathways showing the linkages between a sequence of steps in getting from activities to impact.
- A **theory of change** (ToC) adds to an impact pathway by describing the causal assumptions behind the links in the pathway—what has to happen for the causal linkages to be realized.
- **Causal link assumptions** are the events or conditions necessary or likely necessary for a particular causal link in a ToC pathway to be realized.
- **Results** are the outputs, outcomes, and impacts associated with an intervention.

A discussion of these terms can be found in Mayne (2015). It should be noted that these terms are not always defined or used by others as set out above, and indeed there is no universal agreement on them. It is important, therefore, to define carefully how the terms are being used in a particular setting.

A BRIEF OVERVIEW OF CONTRIBUTION ANALYSIS

Contribution analysis is an approach for addressing causality, producing credible claims about the intervention as a contributory cause (Mayne, 2011, 2012). As such, it explores how and why interventions are working and for whom. Contribution analysis is increasingly being used in evaluations (Buckley, 2016; Buregeya, Brousselle, Nour, & Loignon, 2017; Delahais & Toulemonde, 2017; Downes, Novicki, & Howard, 2018; Kane, Levine, Orians, & Reinelt, 2017; Noltze, Gaisbauer, Schwedersky, & Krapp, 2014; Ton, 2017), and in particular to address causal issues in complex settings (Koleros & Mayne, 2019; Palladium, 2015).

The basis of the contribution claim is the empirical evidence confirming a solid ToC of an intervention, that is, confirming the impact pathways, the assumptions behind the causal links in the ToC, and the related causal narratives explaining how causality is inferred. The ToC is the outline for the *contribution story* of the intervention. The steps usually undertaken in contribution analysis are shown in Box 1 (Mayne, 2011).

Box 1. Steps in contribution analysis

- Step 1: Set out the specific cause-effect questions to be addressed
- Step 2: Develop robust theories of change for the intervention and its pathways
- Step 3: Gather the existing evidence on the components of the theory of change model of causality:
 - The results achieved
 - The causal link assumptions realized
- Step 4: Assemble and assess the resulting contribution claim, and the challenges to it
- Step 5: Seek out additional evidence to strengthen the contribution claim
- Step 6: Revise and strengthen the contribution claim
- Step 7: Return to Step 4 if necessary

KEY CONCEPTS IN CAUSALITY AND CONTRIBUTION ANALYSIS

Causality is always a key element of an evaluation, and hence what perspective to take on causality is important. Contribution analysis—and other theory-based evaluation approaches—uses a generative view of causality, talking of causal packages and contributory causes.

Generative Causality

In many situations a counterfactual perspective on causality—which is the traditional evaluation perspective—is unlikely to be useful; experimental designs are

often neither feasible nor practical. Rather, a more useful perspective is that of *generative causality*:¹ seeing causality as a chain of cause-effect events (Gates & Dyson, 2017, p. 36; Pawson & Tilley, 1997). This is what we see with models of interventions: a series or several series of causal steps—impact pathways—between the activities of the intervention and the desired impacts. Taking the generative or this *stepwise* perspective on causality and setting out an *impact* or *contribution pathway* is essential in understanding and addressing the contribution made by the intervention. The associated ToC model sets out what is needed if the expected results are to be realized.

Contributory Causes

Contribution analysis aims at arriving at credible claims on the intervention as a *contributory cause*, namely, that the intervention was one of several necessary or likely necessary² factors in a *causal package* that together brought about or contributed to the changes observed (Cartwright & Hardie, 2012; Mackie, 1974; Mayne, 2012). That is, it is this *causal package* of factors that will bring about change, and all of these factors are necessary to bring about the change—they are all INUS conditions³—and hence in a logical sense all are of equal importance. In more complex settings, interventions may comprise a number of different components, and for each, one can ask if the component was a contributory cause.

Contribution analysis uses this stepwise perspective on causality to assess whether the intervention has “made a difference,” which in this context means that the intervention had a positive impact on people’s lives—that is, it made a contribution, it played a *causal role*. And it did so because it was a necessary part of a causal package that brought about or contributed to change. This interpretation of making a difference needs to be distinguished from the meaning associated with the counterfactual perspective on causality, where “made a difference” means “what would have happened without the intervention.” This concept of a contributory cause responds to the question posed by Budhwani and McDavid (2017) on the specific meaning of a contribution within CA.

Contribution Claims

Contribution claims have been discussed in previous articles (Mayne, 2011, 2012). But some elaboration and extension is needed. Contribution claims are not just about whether the intervention made a contribution or not. Certainly, a key contribution claim is the yes/no evaluation question:

1. Has the intervention (or component) made a difference? Has it played a positive causal role in bringing about change?

But a more interesting and important contribution claim is around this evaluation question:

2. How and why has the intervention (or component) made a difference, or not, and for whom?

The contribution claim here is about the intervention (or an intervention component) causal package at work. How and in what manner did the intervention support factors and the intervention efforts bring about, or contribute to, change? The contribution claim provides the evidence on why change occurred, that is, the causal narrative. It might also explain why the expected change was *not* realized, why the intervention did not make a difference.

Demonstrating Contribution Claims

As noted above, the basis for contribution analysis is the intervention ToC, and verifying the ToC—the results, the assumptions, and the causal links—with empirical evidence.

Several authors have suggested that in contribution analysis, contribution claims are indeed based on opinions. [Schmitt and Beach \(2015, p. 436\)](#) claim that “[i]n CA, stakeholders [being] interviewed to find out whether they believe the program worked” is the basis for contribution claims. However, this is not what CA is about at all. The aim of contribution analysis is to get beyond basing a contribution claim on opinions of stakeholders about the contribution made. Interviews may be conducted as part of the process to gather information on the results achieved and if assumptions were realized, but basing contribution claims on opinions about the claims is not part of the process. Rather, the evidence gathered on the ToC is used to analyze and make conclusions about contribution claims. Any reports or articles that rely solely on opinions are not reporting on a CA, despite what their authors may claim. Such studies should have a different label to remove references to actual CA.

A second issue related to contribution claims focuses on the role of external factors in arriving at a credible contribution claim. There is indeed some confusion over the role of external influences and especially alternative or rival explanations in CA, confusion that I have contributed to. In [Mayne \(2011\)](#), I suggested that a contribution claim could be made when external factors were shown not to have contributed significantly, and in [Mayne \(2012\)](#), I raised the need to explore rival explanations. These statements were incorrect in that they did not fully recognize the implication of having multiple causal factors at work, some of which may be associated with the intervention and others with external influences. However, external causal factors are usually not alternative or rival explanations. They are simply other causal factors at work.

Therefore, in my view, the “alternative” and “rival” terms are inappropriate in the context of complex causality. But there is a more important implication, namely, that one can explore whether or not a causal factor in a causal package made a contribution and how it did so without considering the other causal factors at play, outside the package, such as external influences, except of course if they are causally linked. A robust ToC sets out the intervention as a contributory cause. Empirically verifying the ToC allows the contribution claim to be made.

Budhwani and McDavid (2017, p. 4) write that “[CA] relies on tests of alternative explanations to act as substitute candidates in place of counterfactuals to determine the plausibility of a proposed theory of change.” As discussed above, this is not the case, and of course CA uses a stepwise (generative) not a counterfactual approach to causality. Lemire et al. (2012) also argue for the need to examine alternative or rival explanations to prove plausible association. Again, this is not correct, but in this case the authors seem to realize this in a footnote, saying that examining alternative explanations is only needed if the aim is to compare the relative contribution of the intervention. And that is true, although I would still argue that the alternative/rival explanations terminology is misleading, since all such factors may be contributing to the results: they are not rivals or alternatives.

The extent to which an evaluation explores the causal factors other than the intervention depends of course on the evaluation question being addressed. If the evaluation question is about assessing what brought about an observed impact, then these other factors would indeed need to be explored. If addressing the narrower question of whether the intervention made a contribution to the impact and how it did so, then these other factors need not play a major role in the analysis (Befani & Mayne, 2014).

If an analysis uses a weak ToC with insufficient causal link assumptions, then a credible contribution claim based on this ToC is not possible. In this case, exploring other external influences might allow some conclusions to be reached concerning the intervention; however, this approach is not CA as discussed in this article.

MEANINGFUL CONTRIBUTION QUESTIONS

Step 1 in contribution analysis (Box 1) is setting out the causal questions to be addressed in the analysis. This is an important first step that is often not adequately addressed. The challenge here is that it is relatively easy to set out evaluation causal questions that sound reasonable and meaningful—such as “Has the intervention been effective?”—but are actually not. The basic reason is that most interventions *on their own* are not the cause of observed results (Mayne, forthcoming). The focus in CA is on the *contribution* an intervention is making to an expected result. Thus, (1) the particular result(s) of interest need(s) to be clearly specified, and (2) CA is not trying to explain what brought about the result, but rather if and how the intervention made a contribution. Therefore, for example, as discussed above, the need to explore other influencing factors depends on just what the causal question is.

USEFUL THEORIES OF CHANGE

The Need in CA for Robust ToCs

Previous articles (Mayne, 2001, 2011, 2012) on contribution analysis generally assume that the ToC used is reasonably detailed and sound, although they do not elaborate. However, *using a weak ToC in a contribution analysis can only lead to weak contribution analysis findings.*

I have suggested criteria for robust theories of change, based on the ToC being both structurally sound and plausible. The detailed criteria, drawn in part from [Davies \(2012\)](#), are discussed in [Mayne \(2017\)](#) for all elements of a ToC: each result, each assumption, each causal link, and overall. For example, if the ToC is not understandable, the causal links in the model cannot be confirmed or, if seemingly “confirmed,” would not lead to credible causal claims. Similarly, if terms are ambiguous, the specific results cannot be empirically confirmed.

As a result of this expanded thinking, Step 2 in [Box 1](#) now highlights the need for a *robust* theory of change. However, the full set of the robust criteria is quite demanding, and the aim is often to ensure that a reasonably robust ToC is available for contribution analysis. The proposed criteria can support this analysis and help strengthen the ToC. In addition, a good ToC should be supported as much as possible by prior social science research and evidence. This type of support will help build credible causal narratives.

Both [Budhwani and McDavid \(2017\)](#) and [Delahais and Toulemonde \(2012\)](#) raise concerns about bias in arriving at contribution claims. I would argue that if one is using a reasonably robust ToC and empirically confirming it in a CA, then the likelihood of bias is greatly reduced, when all of the necessary assumptions associated with each causal link in the ToC are confirmed with empirical evidence. And, of course, if, as [Delahais and Toulemonde \(2012\)](#) argue, one is able to use more than one source of data for the verification, then the chance of any bias is even further reduced. Remember that one is not simply looking to confirm a yes/no issue of contribution but probably, more importantly, from the collection of verified assumptions building a credible causal narrative on how and why the intervention contributed, and for whom.

Some have questioned the need for the “necessity” of causal link assumptions—a robust criterion, noting, in particular, that assumptions are often not 0–1 variables but stated as conditions that could be partially met. What then does necessity mean for a partially met assumption? Results in most ToCs are not defined as a specific amount of the result. Consider an intervention aimed at educating mothers about the benefits of a nutritious diet for their children (see [White, 2009](#), for a discussion of such an intervention). One result here would be “mothers adopt new feeding practices,” and a related assumption could be “supportive husbands and mother-in-law.” Then a partially met assumption (somewhat supportive) would mean less of the result (adopting some practices) but one that is still necessary to get that result.

For a robust ToC, it is always better to define the result as clearly as possible, such as, for example, “fully adopting new practices” to relate better to the causal link assumptions. It is still the case that if the assumption is not realized at all, then there will be no result.

Unpacking Complex Intervention Settings: Different ToCs for Different Purposes

It should be evident that there is not a unique representation of a theory of change for a given intervention, so deciding on how much detail to include can

be a challenge. In most cases, several different depictions of a theory of change are needed to meet different purposes (Mayne, 2015). Further, ToCs can quickly become overly complicated and less useful if too much detail is used in any one representation. In Mayne (2015), several levels of ToCs are presented and their uses discussed to help with this problem:

- *A narrative ToC* describes briefly how the intervention is intended to work.
- *The overview ToC* indicates the various pathways to impact that comprise the intervention showing some of steps in each pathway along the way to impact. It can also set out the rationale assumptions or premises behind the intervention, but usually not the causal link assumptions.
- *Nested ToCs* are developed to unpack a more complicated/complex intervention and include the explicit causal link assumptions. There can be a nested ToC, for example, for each pathway, for each pathway in a different geographical area, and/or for different targeted reach groups. Koleros and Mayne (2019) discuss using nested ToCs for different actor groups for a complex police reform intervention.

Budhwani and McDavid (2017, p. 19) suggest that CA may not work well in complex settings due to the difficulty of building useful ToCs in such a context. Actual experience is quite the opposite. Using nested ToCs to unpack a complex intervention and its context has worked well in numerous situations (see, for example, Douthwaite, Mayne, McDougall, & Paz-Ybarnegaray (2017); Koleros & Mayne (2019); Mayne & Johnson (2015); Riley et al. (2018).

The Need for Evaluable ToC Models

Usually, the evaluator needs to develop or guide the development of ToCs that can be used for evaluation purposes. Often, the evaluator finds an already developed ToC of the intervention being evaluated, but it may not be suitable for evaluation purposes (for instance, it may be well suited for acquiring funding or communication purposes). Something more evaluable is needed, such as developing nested ToCs to unpack the complexity of the intervention, with careful thought given to the causal assumptions at play.

Developing “good” ToCs is itself a challenge, but equally it is often a serious challenge to bring on board those who “own” the existing ToC and may not want to see a new ToC brought into play. Koleros and Mayne (2019) discuss handling this situation.

Behaviour-Change ToC Models

Most interventions involve changing the behaviour of one or more actor groups, so behaviour change needs to be a key focus (Earl, Carden, & Smutylo, 2001). The detailed ToCs needed for CA can be based on the generic behaviour-change ToC model, shown in Figure 1 (Mayne, 2017, 2018). The model is a synthesis of social

science research on behaviour change by [Michie, Atkins, and West \(2014\)](#), which argues that behaviour (B) is changed through the interaction of three necessary elements: capabilities (C), opportunities (O), and motivation (M). Hence the name: the COM-B model.

The COM-B ToC model has proven very useful for building robust nested ToCs and for undertaking contribution analysis, because it is quite intuitive and is based on a synthesis of empirical evidence on behaviour change. It is especially helpful in explaining how behaviour changes were brought about. That is, the COM-B model is a model of the mechanisms⁴ at work in bringing about behaviour change and thus provides the basis for inferring causality about behaviour change.

A number of authors who have used contribution analysis in complex settings have noted, though, that it can be quite data- and analysis-demanding, when one has to work with a large number of nested ToCs ([Delahais & Toulemonde, 2012](#); [Freeman, Mayhew, Matsinhe, & Nazza, 2014](#); [Noltze et al., 2014](#); [Schmitt and Beach \(2015\)](#) make a similar note regarding process tracing, which is closely related to CA ([Befani & Mayne, 2014](#)).

Figure 1. The COM-B Theory of Change model

With that in mind, intermediate level ToCs would be useful—more than the overview ToC but less detailed than an operational nested ToC. This is where a *simplified ToC* could be useful. The idea of a simplified ToC is to develop a less complex ToC in the context of a contribution analysis, especially when there may be quite a few pathways to analyze. So, for example, rather than the more detailed generic behaviour-change ToC (Figure 1), we might have, more simply, activities/outputs, behaviour change, direct benefits, and impact (Figure 2) as the pathway ToC. Figure 2 shows the essence of the pathway getting from activities/outputs to impact, explicitly identifying results that are usually straightforward to measure. The associated causal link assumptions would normally include the following assumptions:

- the intended target groups were reached, and
- adequate improvements in capabilities, opportunities, and motivation were achieved.

In setting out the causal link assumptions, a detailed nested ToC for the pathway is almost essential for their identification. The aim would be to have a minimum number of higher-level assumptions in the simplified ToC, perhaps arrived at by aggregating assumptions from the more detailed ToC.

Figure 2 shows one model for a simplified ToC. Even simpler pathways could be developed, such as dropping the behaviour change box, or the direct benefit

Figure 2. A simplified COM-B Theory of Change

box. Then the pathway assumptions would have to include behaviour change and direct benefits, respectively, in order to keep those in the model.

Simplified ToCs would reduce the amount of data required to carry out a contribution analysis to determine if the intervention had made a contribution or not. However, in order to understand why the intervention did or did not work, one would need to focus on the behaviour-change level. But determining why the expected behaviour changes did not come about, for instance, can be done retrospectively, asking those involved about reach and capacity change (capabilities, opportunities, and motivation).

Experience to date does suggest the need to first develop a detailed nested ToC, and then the simplified version. In this way it becomes clear what is being suppressed in the simplified ToC and needs to be kept in mind, even though the simplified ToC would actually be used in Contribution Analysis.

HOW MUCH OF A CONTRIBUTION?

There remains in CA a desire to say something about the quantitative size of the contribution a causal factor is making. [Budhwani and McDavid \(2017, p. 20\)](#) talk about measuring the degree of contribution so that the CA can reach findings similar to cost-benefit analysis. This is not possible because of the nature of complex causality. There are multiple causal factors at work, and it is packages of necessary causal factors that bring about change, not any individual factor. Although others have attempted to examine the issue of estimating size effects within contribution analysis ([Ton et al., 2019](#)), CA does not, on its own, estimate the size or indeed the relative importance of the causal factors at work.

But exploring the relative importance question is possible ([Mayne, 2019](#)). There is a need, then, to carefully decide (a) which causal factors one wants to compare and (b) how one wants to interpret “importance.” A variety of perspectives are possible: perceived importance, the roles played by the factors, the funds expended, and the extent of the constraints to change. All are plausible ways of assessing the relative importance of causal factors.

INFERRING CAUSALITY: DEMONSTRATING CONTRIBUTION CLAIMS

CA aims to result in claims about the contribution made by an intervention to observed results. A first question, then, is which results? In looking at an intervention and its ToC, it is clear that there could be a number of interesting contribution claims, namely, claims associated with any of the results along the impact pathway. Contribution claims for early results would probably be quite easily established, while more distant outcomes and impact are likely to present more of a challenge. But it would be important to identify just which contribution claims were of prime interest.

And of course, claims for more distant results need to be built on credible claims for the earlier pathway results. Hence the need to consider approaches to verifying a single causal link in a ToC. In a more complex intervention there would be several different pathways to impact, each with its own ToC. And often, it is useful to know if each of these pathways contributed to the success (or not) of the intervention. For example, in the case where actor-based ToCs have been developed for the intervention, it is of considerable interest to understand how and why the various actor groups contributed to bring about results.

Causal Inference Analysis

Key to credible contribution claims are credible arguments inferring causality—the logic and evidence used to justify a causal link—which would be used in Step 4 to assess the contribution story to date. An evidence-based contribution claim has two parts:

1. The intervention (or a component) contributed to an observed change—it played a positive role in bringing about change, and
2. It did so in the following manner ...

Showing that the intervention was a contributory cause accomplishes both of these aims: the intervention is part of a causal package that was sufficient to bring about the change—which explains how the change was brought about (2), and that the intervention was a necessary part of the causal package (1), and hence a causal factor in bringing about the change. Process tracing is a useful alternative way for getting at (1), but it does not provide the information needed for (2).

[Befani and Mayne \(2014\)](#) and [Befani and Stedman-Bryce \(2016\)](#) have noted correctly that while CA seeks to verify the causal links in a theory of change, previous discussions ([Mayne, 2011, 2012](#)) do not say much about how to go about doing the verification. Yet this is a key step and more of a challenge when examining complex interventions. This article looks more closely at making these causal inferences and builds on the approach of process tracing and related insights on causality, arguing the need for solid *causal narratives*.

In the traditional CA approach, showing that the intervention was a contributory cause and hence made a difference—that is, contributed to an observed impact and how it did so—requires demonstrating that

- the theory of change (the causal package) was sufficient, and
- the intervention activities were an essential part of the causal package, and hence a causal factor in bringing about change.

Sufficiency is demonstrated by showing that each causal link in the theory of change (ToC) with its assumptions was realized. Sufficiency was always a weak point in the argument, and I would now say that data showing the ToC was realized is not enough. One needs in addition to build credible causal narratives for the ToC. This picks up a key point made by [Pearl and Mackenzie \(2018\)](#) in their

The Book of Why, namely that statistics alone are not enough to infer causality; one also needs good explanatory causal theory. As mentioned previously, good ToCs are often based on some social science theory and not just the thoughts of a program team, so that they can provide the basis for solid causal explanations. What is needed is good causal reasoning (Davidson, 2013).

Let me first note again that CA is expected to be done on a reasonably robust ToC, and many of the criteria for robustness are indeed criteria for inferring causality, forming the elements of a credible causal narrative. Table 1 sets out four tools for inferring causality, all of which are embedded in a robust ToC and described in more detail below.

The evidence tools in Table 1 can be used to build credible causal narratives. Causal narratives provide the argument and evidence related to how the causal factors at work played a positive role in bringing about change. They explain the how a causal link worked, or the causal mechanisms at play. In Table 1, the “Robust ToC #” values are references to the robust criteria in Mayne (2017).

Causal Inference Evidence Tools

Checking that Change Occurred

1. *Verifying the ToC.* With a robust ToC, verifying that the pathway results and associated assumptions were realized lays the basis for the plausibility of a contribution claim. As Weiss (1995, p. 72) argues, “Tracking the micro-stages of the effects as they evolve makes it more plausible that the results are due to program activities and not to outside events or artifacts of the evaluation, and that the results generalize to other programs of the same type.” Verifying the ToC provides the empirical evidence on which causal narratives are built. If aspects of the ToC cannot be verified, then causal claims cannot be made about those aspects.

The next three tools are hoop tests used in process tracing. If the verified ToC does not reflect them, then causality is unlikely. However, confirming these three tests does not confirm causality, as there may be other causal factors at work.

Hoop Tests for Confirming Plausibility

2. *Logical and plausible time sequence of results and assumptions.* The evidence sought here is that

- the results along a pathway were realized in a logical time sequence (i.e., cause preceded effect along the causal chain);
- the assumptions for each causal link were realized after the preceding result, i.e., were pre-events and conditions for the subsequent result; and
- the timing of when the results were realized was plausible and consistent with the ToC timeline.

This may seem like an obvious criterion, but in practice it can prove quite useful. Too often, for example, ToCs do not have a timeline and hence the third component of the criterion cannot be applied. It can easily be the case that a result

Table 1. Evidence for inferring causality

Tools	References	Comment
<i>Checking that change occurred</i>		
1. Verifying pathway and assumptions, including at-risk assumptions	<ul style="list-style-type: none"> • Robust ToC #9 • Contribution Analysis • Weiss (1995) 	Are the pathway and assumptions verified? This forms the evidence base for making the contribution claims.
<i>Hoop tests for confirming plausibility</i>		
2. Logic and plausible time sequence	<ul style="list-style-type: none"> • Robust ToC #3 : timing • Davidson (2009) • Robust ToC #4: Logical coherence 	<p>Needed to explain causality.</p> <p>Link: Are assumptions pre-events and conditions for the result?</p> <p>ToC: Is sequence of results plausible?</p> <p>Is the timing of the occurrence of the results plausible?</p>
3. Reasonable effort expended	<ul style="list-style-type: none"> • Robust ToC #11: level of effort • Davidson (2009) 	Is it reasonable that the level of effort expended will deliver the results?
4. Expect-to-see effects realized	<ul style="list-style-type: none"> • Process tracing: hoop test 	If effects not seen, causality very unlikely. But effects might have other causes.
<i>Building the causal narrative</i>		
5. Causal packages are sufficient	<ul style="list-style-type: none"> • Robust ToC #10: A sufficient set • Robust ToC #5: necessary or likely necessary assumptions 	<p>Is it reasonable that the collection of causal package factors is sufficient to bring about the result?</p> <p>Are the mechanisms at work identified?</p> <p>Have the barriers to change been addressed?</p>
<i>Confirming a causal factor</i>		
6. Some unique effects observed	<ul style="list-style-type: none"> • Process tracing: smoking gun tests 	Result only possible if intervention is the cause.

was not realized because not enough time has elapsed. Conversely, if a result has indeed appeared but earlier than expected, it may suggest something other than the intervention at work. Furthermore, confirming that the assumptions were realized in a timely fashion means that the basis for the causal narrative for the link is sound. Taken together, the three points above argue that the causal link is quite plausible.

3. *Reasonable effort expended.* Again this is a plausibility test. If, in implementation, the size of the actual intervention, including the efforts of any partners,

appears quite small in comparison to the observed results, then a contribution claim may seem implausible (Davidson, 2009).

4. *Expected-to-see effects realized.* This is the process-tracing hoop test (Punton & Welle, 2015) whereby if the causal link has worked, then there are effects, often secondary effects, that one would expect to see. If those effects are not realized, then the causal link is doubtful.

Building the Causal Narrative

5. *Causal packages are sufficient.* This is the essential tool in building the causal narrative. We are trying to build an argument that the causal link between one result (R1) along an impact pathway—the cause—and its subsequent result (R2)—the effect—worked. We would have shown that R1 and R2 did occur, as did the associated causal link assumptions. The set of assumptions in particular set out the framework for the argument, for the causal “story.” In bringing about R2, one can imagine various constraints or barriers to change. The assumptions are events and conditions that are expected to overcome these barriers. This can be a useful way to develop the causal narrative.

A related approach is using causal mechanisms. Realist evaluation (Westhorp, 2014) argues that causality can be inferred by identifying and explaining the causal mechanisms at work and the contexts in which the intervention occurs. In a ToC approach, the context and the mechanisms at work are captured by the causal link assumptions.

Schmitt and Beach (2015) have claimed that ToCs “hide” the mechanism at work. While the realist causal mechanisms are not explicit in many ToC models and hence CA, CA uses a different paradigm to conceptualize causality, namely causal packages. Further, the causal mechanisms can often be readily identified by working through the causal package at work. Delahais and Toulemonde (2017, p. 385), in discussing their contribution analysis work, make this link:

In the process of translating the “framing pathway” into a “framing mechanism,” we may consider that we have just refined the description of the causal package, i.e. deepened the exercise without changing its nature. We have often had this impression while reading illustrations of the concept of mechanism In fact the very change in the nature of the exercise occurs when the mechanism is given a name and referred to the literature, i.e. when we assume that it remains the same in different contexts and then acquires its generalization potential.

That is, the advantage of using causal mechanisms is that they refer to more general causal forces at work, as referred to in the literature, and hence provide common-sense logical explanations of causality. Let me note in particular that the social science research-based COM-B ToC model explicitly identifies the causal mechanisms at work, namely capability, opportunity, and motivation to bring about behaviour change.

The bottom line is to set out a sound and valid argument—a causal narrative—of why the causal package at work did indeed contribute to R2.

Confirming a Causal Factor

6. *Unique effects realized.* This is the process-tracing smoking gun test for a causal factor. Unique effects with respect to a specific causal factor are effects that can be realized only if the causal factor was indeed part of the causal package bringing about change (Befani & Mayne, 2014; Punton & Barnett, 2018). If they are observed, then this is strong evidence that the causal factor played a positive causal role in bringing about R2. But note that this test does not provide evidence of how the change was brought about, that is, what the other factors in the causal package are.

GENERALIZING CA FINDINGS

Contribution analysis shows that an intervention in a specific location contributed to an observed result and how it did so. What might be said about the intervention implemented in a different location? This is the issue of external validity or generalization of CA findings.

If the intervention ToC worked in the new location, would it play the same positive causal role there? To conclude this, one would need to show that it was likely that in the new location

- the intervention could deliver the same (or quite similar) outputs,
- the causal link assumption would be realized, and
- the causal narratives would remain valid.

The likelihood of each of these could be assessed. To the extent that higher-level causal assumptions have been used in the ToC, such as when causal mechanisms have been identified, then the argument that the causal narratives remain valid will be stronger. In the nutrition example mentioned earlier, a key assumption needed was that mothers control food distribution in the household (an assumption that was missed initially). However, the more general causal assumption is that there is a need to education the person(s) in power in the household—which might not be the mother—a higher-level assumption.

One would need to carefully assess the conditions outlined above to produce a finding about the generalizability of an intervention. Cartwright and Hardie (2012, p. 7) argue that generalizing follows if, in a new location, the intervention played the same causal role and the support factors (the causal link assumptions) are in place. This is the same rationale as the CA argument above, using slightly different terms.

Clearly, if there is something very unique about the original location in some causal link assumptions, then generalizing is unlikely to be possible.

SO WHITHER CONTRIBUTION ANALYSIS?

Contribution analysis was set out some years ago as a set of general steps to take in addressing causality. As such, over a number of years it led to a variety of ways

of operationalizing the concepts and principles, with numerous suggestions being made for applying CA in specific cases. This has all been for the good. There have also been numerous articles raising legitimate questions about CA and its application. In this article I have tried to look back at how CA has been applied and consider the concerns that have been expressed.

In the last few years, I have seen a significant rise in applications of CA, particularly as applied to complex settings, which are becoming more common. And indeed, given that it assumes that multiple causal factors and interventions can play a contributory role, it can be well suited to address causality in those settings, especially using nested ToCs. I expect to see more and more applications of CA in a variety of settings. But there is a need to be clear about what contribution analysis can and cannot do.

Contribution analysis is not a quick-and-dirty approach to addressing causality. On the downside, (1) it often does require a substantial amount of data, along with rigorous thinking, (2) it requires reasonably robust theories of change, and (3) it cannot determine how much of an outcome result can be attributed to an intervention.

On the other hand, it offers several advantages: (1) it can be used to make causal inferences when experimental and quasi-experimental designs are not possible, or not needed/desired; (2) it explores why and how an intervention has influenced change, and for whom; (3) it can be part of a mixed-method approach to an evaluation, such as when using comparative groups to assess how much change has occurred; (4) it allows for making causal inferences about the intervention without necessarily examining external causal factors; and (5) it addresses cases where there are numerous causal factors at work by assessing contributory causes leading to credible contribution claims.

Overall, CA has been found to be a practical way to explore causal relationships and to better understand how changes have been brought about, and for whom.

NOTES

- 1 For a discussion on different perspectives on causality, see Befani's Appendix in [Stern et al. \(2012\)](#).
- 2 "Likely necessary" allows for a probabilistic interpretation of an assumption ([Mahoney, 2008](#), p. 421). See [Mayne \(2015, p. 126\)](#) for a discussion.
- 3 That is, they are an Insufficient but Necessary part of a condition that is itself Unnecessary but Sufficient for the occurrence of the effect ([Mackie, 1974](#)). See [Mayne, \(2012, p. 276\)](#) for a discussion of these INUS conditions.
- 4 Realist evaluations use the concept of mechanisms to infer causality ([Pawson & Tilley, 1997](#); [Westthorp, 2014](#)).

REFERENCES

- [Befani, B., & Mayne, J. \(2014\)](#). Process tracing and contribution analysis: A combined approach to generative causal inference for impact evaluation. *IDS Bulletin*, 45(6), 17–36. <https://doi.org/10.1111/1759-5436.12110>

- Befani, B., & Stedman-Bryce, G. (2016). Process tracing and Bayesian updating for impact evaluation. *Evaluation*, 23(1), 42–60. <https://doi.org/10.1177/1356389016654584>
- Buckley, A. P. (2016). Using Contribution Analysis to evaluate small & medium enterprise support policy. *Evaluation*, 22(2), 129–148. <https://doi.org/10.1177/1356389016638625>
- Budhwani, S., & McDavid, J. C. (2017). Contribution analysis: Theoretical and practical challenges and prospects for evaluators. *Canadian Journal of Program Evaluation*, 32(1), 1–24. <https://doi.org/10.3138/cjpe.31121>
- Buregeya, J. M., Brousselle, A., Nour, K., & Loignon, C. (2017). Comment évaluer les effets des évaluations d'impact sur la santé : le potentiel de l'analyse de contribution. *Canadian Journal of Program Evaluation*, 32(1), 25–45. <https://doi.org/10.3138/cjpe.31151>
- Byrne, D. (2013). Evaluating complex social interventions in a complex world. *Evaluation*, 19(3), 217–228. <https://doi.org/10.1177/1356389013495617>
- Cartwright, N., & Hardie, J. (2012). *Evidence-based policy: Doing it better. A practical guide to predicting if a policy will work for you*. Oxford, England: Oxford University Press.
- Davidson, E. J. (2009). *Causation inference: Nuts and bolts*. A Mini Workshop for the ANZEA Wellington branch. Wellington, New Zealand. Retrieved from <http://realevaluation.com/pres/causation-anzea09.pdf>
- Davidson, E. J. (2013). *Understanding causes and outcomes of impacts*. BetterEvaluation coffee break webinars. Retrieved from https://www.betterevaluation.org/events/coffee_break_webinars_2013#webinarPart5
- Davies, R. (2012). Criteria for assessing the evaluability of Theories of Change. *Rick on the road* [blog]. Retrieved from <http://mandenews.blogspot.com/2012/04/criteria-for-assessing-evaluability-of.html>
- Delahais, T., & Toulemonde, J. (2012). Applying contribution analysis: Lessons from five years of real life experience. *Evaluation*, 18(3), 281–293. <https://doi.org/10.1177/1356389012450810>
- Delahais, T., & Toulemonde, J. (2017). Making rigorous causal claims in a real-life context: Has research contributed to sustainable forest management? *Evaluation*, 23(4), 370–388. <https://doi.org/10.1177/1356389017733211>
- Douthwaite, B., Mayne, J., McDougall, C., & Paz-Ybarnegaray, R. (2017). Evaluating complex interventions: A theory-driven realist-informed approach. *Evaluation*, 23(3), 294–311. <https://doi.org/10.1177/1356389017714382>
- Downes, A., Novicki, E., & Howard, J. (2018). Using the contribution analysis approach to evaluate science impact: A case study of the national institute for occupational safety and health. *American Journal of Evaluation*, 40(2), 177–189. <https://doi.org/10.1177/1098214018767046>. Medline:30518992
- Dybdal, L., Nielsen, S. B., & Lemire, S. (2010). Contribution analysis applied: Reflections on scope and method. *Canadian Journal of Program Evaluation*, 25(2), 29–57.
- Earl, S., Carden, F., & Smutylo, T. (2001). *Outcome mapping*. Ottawa, ON: International Development Research Centre.
- Freeman, T., Mayhew, S., Matsinhe, C., & Nazza, D. A. (2014). *Evaluation of the Danish strategy for the promotion of sexual and reproductive health and rights 2006–2013—Pathways to change in SRHR: Synthesis report*. Ministry of Foreign Affairs of Denmark.

- Retrieved from http://www.netpublikationer.dk/um/14_evaluation_2014_03/Pdf/evaluation_2014_03.pdf
- Gates, E., & Dyson, L. (2017). Implications of the changing conversation about causality for evaluators. *American Journal of Evaluation*, 38(1), 29–46. <https://doi.org/10.1177/1098214016644068>
- Gerrits, L., & Verweij, S. (2015). Taking stock of complexity in evaluation: A discussion of three recent publications. *Evaluation*, 21(4), 481–491. <https://doi.org/10.1177/1356389015605204>
- Kane, R., Levine, C., Orians, C., & Reinelt, C. (2017). *Contribution analysis in policy work: Assessing advocacy's influence*. Centre for Evaluation Innovation. Retrieved from https://www.evaluationinnovation.org/wp-content/uploads/2017/11/Contribution-Analysis_0.pdf
- Koleros, A., & Mayne, J. (2019). Using actor-based theories of change to conduct robust contribution analysis in complex settings. *Canadian Journal of Program Evaluation*, 33(3), 292–315. <https://doi.org/10.3138/cjpe.52946>
- Lemire, S. T., Nielsen, S. B., & Dybdal, L. (2012). Making contribution analysis work: A practical framework for handling influencing factors and alternative explanations. *Evaluation*, 18(3), 294–309. <https://doi.org/10.1177/1356389012450654>
- Mackie, J. L. (1974). *The cement of the universe: A study of causation*. Oxford, England: Oxford University Press.
- Mahoney, J. (2008). Toward a unified theory of causality. *Comparative Political Studies*, 41(4/5), 412–436. <https://doi.org/10.1177/0010414007313115>
- Mayne, J. (2001). Addressing attribution through contribution analysis: Using performance measures sensibly. *Canadian Journal of Program Evaluation*, 16(1), 1–24. Retrieved from <https://evaluationcanada.ca/system/files/cjpe-entries/16-1-001.pdf>
- Mayne, J. (2011). Contribution analysis: Addressing cause and effect. In R. Schwartz, K. Forss, & M. Marra (Eds.), *Evaluating the complex* (pp. 53–96). New Brunswick, NJ: Transaction Publishers.
- Mayne, J. (2012). Contribution analysis: Coming of age? *Evaluation*, 18(3), 270–280. <https://doi.org/10.1177/1356389012451663>
- Mayne, J. (2015). Useful theory of change models. *Canadian Journal of Program Evaluation*, 30(2), 119–142. <https://doi.org/10.3138/cjpe.230>
- Mayne, J. (2017). Theory of change analysis: Building robust theories of change. *Canadian Journal of Program Evaluation*, 32(2), 155–173. <https://doi.org/10.3138/cjpe.31122>
- Mayne, J. (2018). *The COM-B theory of change model*. Retrieved from https://www.researchgate.net/publication/323868561_The_COMB_ToC_Model4
- Mayne, J. (2019). Assessing the relative importance of causal factors. CDI Practice Paper 21. Brighton, England: IDS. Retrieved from <https://opendocs.ids.ac.uk/opendocs/handle/123456789/14647>
- Mayne, J. (Forthcoming). Realistic commissioning of impact evaluations: Getting what you ask for? In A. Paulson & M. Palenberg (Eds.), *Evaluation and the pursuit of impact*. Abingdon, England: Taylor and Francis.

- Mayne, J., & Johnson, N. (2015). Using theories of change in the CGIAR Research Program on Agriculture for Nutrition and Health. *Evaluation*, 21(4), 407–428. <https://doi.org/10.1177/1356389015605198>
- Michie, S., Atkins, L., & West, R. (2014). *The behaviour change wheel: A guide to designing interventions*. London, England: Silverback Publishing. Retrieved from <http://www.behaviourchangewheel.com/>
- Noltze, M., Gaisbauer, F., Schwedersky, T., & Krapp, S. (2014). Contribution analysis as an evaluation strategy in the context of a sector-wide approach: Performance-based health financing in Rwanda. *African Evaluation Journal*, 2(1), 8 pp. Retrieved from <http://www.aejonline.org/index.php/aej/article/view/81>
- Palladium. (2015). *Independent evaluation of the Security Sector Accountability and Police Reform Programme: Final evaluation report*. DFID. Retrieved from http://r4d.dfid.gov.uk/pdf/outputs/SSAPRP-DRC/SSAPR_Complete_Impact_Evaluation_Rep_Final-Eng.pdf
- Pawson, R., & Tilley, N. (1997). *Realistic evaluation*. London, England: SAGE.
- Paz-Ybarnegaray, R., & Douthwaite, B. (2016). Outcome evidencing: A method for enabling and evaluating program intervention in complex systems. *American Journal of Evaluation*, 38(2), 275–293. <https://doi.org/10.1177/1098214016676573>
- Pearl, J., & Mackenzie, D. (2018). *The book of why: The new science of cause and effect*. New York, NY: Basic Books.
- Punton, M., & Barnett, C. (2018). *Contribution analysis and Bayesian confidence updating: A brief introduction*. Itad briefing paper. Retrieved from <https://itad.com/wp-content/uploads/2019/06/CDC-Brief-contribution-analysis-and-Bayesian-Updating.pdf>
- Punton, M., & Welle, K. (2015). *Straws-in-the-wind, hoops and smoking guns: What can process tracing offer to impact evaluation?* CDI Practice Paper, No. 10. Centre for Development Impact. Retrieved from http://opendocs.ids.ac.uk/opendocs/bitstream/handle/123456789/5997/CDIPracticePaper_10.pdf?sequence=1
- Riley, B. L., Kernaghan, A., Stockton, L., Montague, S., Yessis, J., & Willis, C. D. (2018). Using contribution analysis to evaluate the impacts of research on policy: Getting to “good enough.” *Research Evaluation*, 27(1), 16–27. <https://doi.org/10.1093/reseval/rvx037>. Retrieved from <https://academic.oup.com/rev/article/27/1/16/4554784>
- Schmitt, J., & Beach, D. (2015). The contribution of process tracing to theory-based evaluations of complex aid instruments. *Evaluation*, 21(4), 429–447. <https://doi.org/10.1177/1356389015607739>
- Stern, E., Stame, N., Mayne, J., Forss, K., Davies, R., & Befani, B. (2012). *Broadening the range of designs and methods for impact evaluations*. DFID Working Paper, 38. London, England: DFID. Retrieved from <http://r4d.dfid.gov.uk/Output/189575/>
- Ton, G. (2017). Contribution analysis of a Bolivian innovation grant fund: Mixing methods to verify relevance, efficiency and effectiveness. *Journal of Development Effectiveness*, 9(1), 120–143. <http://doi.org/10.1080/19439342.2016.1231702>
- Ton, G., Mayne, J., Delahais, T., Morell, J., Befani, B., Apgar, M., & O’Flynn, P. (2019). *Contribution analysis and estimating the size of effects: Can we reconcile the possible with the impossible?* CDI Practice Paper, Number 20. Centre for Development Impact.

Retrieved from <https://www.cdimpact.org/publications/contribution-analysis-and-estimating-size-effects-can-we-reconcile-possible-impossible>

- Weiss, C. H. (1995). Nothing as practical as good theory: Exploring theory-based evaluation for comprehensive community initiatives for children and families. In J. P. Connell, A. C. Kubisch, L. B. Schorr, & C. H. Weiss (Eds.), *New approaches to evaluating community initiatives: Volume 1: Concepts, methods and contexts* (pp. 65–92). Washington, DC: The Aspen Institute.
- Westthorp, G. (2014). *Realist impact evaluation: An introduction*. Methods Lab. Retrieved from <http://www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/9138.pdf>
- White, H. (2009). *Theory-based impact evaluation: Principles and practice*. Working Paper 3. International Initiative for Impact Evaluation (3ie). Retrieved from https://www.3ieimpact.org/sites/default/files/2017-11/Working_Paper_3.pdf
- Wilson-Grau, R., & Britt, H. (2012). *Outcome harvesting*. Retrieved from <http://www.outcomemapping.ca/resource/resource.php?id=374>

AUTHOR INFORMATION

John Mayne is an independent advisor on public sector performance. Over the past 13 years he has focused largely on international development evolution and results-based management work.

Self-Evaluation Tool for Action in Partnership: Translation and Cultural Adaptation of the Original Quebec French Tool to Canadian English

Angèle Bilodeau

Université de Montréal

Gillian Kranias

Health Nexus

Abstract: *This article presents the translation and cultural adaptation, into Canadian English, of the Outil diagnostique de l'action en partenariat, a tool widely used to support the practice of partnerships since its creation in French, in Quebec, in 2008. The theoretical foundations and properties of the original tool are presented, followed by a summary of methodological guidelines and a description of the process and results. The methodology involved an expert committee to formulate the English tool and verify its equivalence with the original, and a pretest with target users. This rigorous procedure ensures equivalence of the translated tool and its cultural adaptation to the intended users.*

Keywords: *action in partnership, action network, inter-sectoral action, partnership self-evaluation tool, translation-adaptation of measurement tools*

Résumé : *Cet article présente la traduction et l'adaptation culturelle, en anglais canadien, de l'Outil diagnostique de l'action en partenariat, largement utilisé pour soutenir la pratique des partenariats depuis sa création en français, au Québec, en 2008. Les fondements théoriques et les propriétés de l'outil original sont présentés. Suivent ensuite la démarche méthodologique et la présentation des résultats. La méthodologie comprend un comité d'experts pour la traduction de l'outil vers l'anglais et la vérification de son équivalence avec l'outil original, et un pré-test effectué auprès d'utilisateurs potentiels. Cette procédure rigoureuse assure l'équivalence de l'outil traduit et son adaptation culturelle aux utilisateurs visés.*

Mots clés : *action en partenariat, action en réseau, action intersectorielle, outil d'auto-évaluation des partenariats, traduction-adaptation d'outils de mesure*

Partnership action is a widely recommended strategy for addressing social determinants of complex issues at varying levels of public action (WHO, 2011). The

Corresponding author: Angèle Bilodeau, École de santé publique, Université de Montréal, Montréal, Québec, Canada, 7101 avenue du Parc, Montréal, QC H3C 3J7; angele.bilodeau@umontreal.ca

availability of evidence-based and user-friendly tools to evaluate the quality of partnerships is a significant lever to undertake action in partnership or follow its progress (Fetterman & Wandersman, 2007; Halliday, Asthana, & Richardson, 2004; Newcomer, Hatry, & Wholey, 2015). The *Self-evaluation tool for action in partnership*¹ is a tool widely used in Quebec in its original French version. It was developed in 2008 by researchers from the Montreal Directorate of Public Health and the Canada Research Chair *Approches communautaires et inégalités de santé* of l'Université de Montréal. It met the need for evaluation tools theoretically based on empirical research, was rigorously validated, and was able to fit into the current activities of partnerships. Its Canadian English adaptation meets the needs of practitioners and researchers, responding to a request from Health Nexus in Ontario. This article first details the characteristics of the original tool. Then it summarizes the methodological guidelines for the translation and validation of measurement tools. Finally, it presents the methodology and results of the translation and adaptation of the original tool to the context of English Canada.

THE ORIGINAL TOOL: THE OUTIL DIAGNOSTIQUE DE L'ACTION EN PARTENARIAT

The tool fits within a theory-based evaluation approach (Devaux-Spatarakis, 2014). More specifically, it meets the need for higher-level theories of change—in contrast to local theories—which may result from multi-site analysis, meta-analysis, or the use of general social theories to produce more general interpretive models (Auspos & Kubisch, 2004; Barnes, Matka, & Sullivan, 2003; Chircop, Bassett, & Tatlor, 2015; Dahler-Larsen, 2001; Villeval et al., 2016). The tool operationalizes a mid-range theory that defines six requirements for effective partnership work from a series of case studies based on Actor Network Theory (Bilodeau, Galarneau, Fournier, & Potvin, 2011). It makes it possible to assess the functioning of a partnership, identifying both difficulties and aspects that work well with reference to the six requirements for effectiveness. It is also a self-evaluation tool, allowing practitioners to situate their own partnership according to three levels of achievement for each requirement and to make a judgment founded on empirically established best practices based on case studies.

The first four requirements are related to participation dynamics, that is, setting actors into motion to address a situation that is deemed problematic. (1) *The range of perspectives relevant to the issue*: Partnerships must include the various perspectives on the issue of interest to broaden the opportunities for convergence among actors. (2) *Early stakeholder involvement in strategic decisions*: Partners need to be actively involved in analyzing issues and developing options for action, not just in implementation. (3) *Engagement of stakeholders in negotiating and influencing decisions*: Partners need to be engaged to a degree that goes beyond consultation and must have real influence on decisions. (4) *Commitment of strategic and pivotal stakeholders to the project*: Stakeholders must be in a position to make decisions and commit resources, and essential resources for successful

action must be mobilized. Furthermore, partners must be stable and able to bring in new stakeholders needed to move forward in action together.

The fifth and sixth requirements relate to partnership arrangements. Such arrangements concern structures and mechanisms that build cooperation and support the resolution of controversies, usually linked to questions of power or paradigms of action. (5) *Partnership arrangements that favour equalization of power among the stakeholders*: Favourable arrangements are structures and mechanisms that reduce subordination resulting from structural inequality among actors. Thereby, actors can identify concrete indicators of their influence and interdependence, such as the equitable acknowledgement and distribution of benefits, equal consideration of all points of view in discussion and decision, and negotiated criteria for accountability. (6) *Partnership arrangements that help build collective action*: These are exchanges of diverse points of view that expand possibilities for action, open up discussions over disagreements, and support their successful resolution, going beyond individual interests to find common ground in the interest of the target population and modifying individual actions to build new and more holistic options for shared action. Through such arrangements, actors can concretely identify progress in thinking and action that result from linking partner paradigms and resources.

The tool includes 18 items relating to the six requirements described above. For each item, respondents must choose one of three options representing varying degrees of achievement of the item (strong, moderate, weak) in their partnership. The tool is designed to be used by members of a partnership who voluntarily participate in self-evaluation. Three types of partnership evaluation may be produced: (1) A one-time evaluation can identify the strong/weak items or the strong/weak requirements. The items or requirements that receive the largest number of “strong” responses indicate the strengths of the partnership. Those receiving the largest number of “weak” responses indicate its weaknesses. The items lacking consensus indicate possible areas of controversy. The items left blank indicate that some respondents are unable to assess important aspects of their partnership. (2) Longitudinal evaluation is possible if the exercise is repeated after a period of time. An evaluation performed for each item or each requirement can show improvement or deterioration over time. Finally, (3) a summary portrait can be produced across multiple partnerships within a similar context (e.g., same funder, same issue, or same policy). If more than one partnership is experiencing difficulties with the same items or requirements, this could highlight collective issues.

The original French tool was validated using three methods (Bilodeau et al., 2011). The cognitive interview (Presser et al., 2004) with target user groups served to pre-test the tool in terms of their understanding of items and response choices (first method). This is a participatory pretest method of a questionnaire where participants verbalize their understanding of the questions and answer choices. Two group interviews of five and six participants were conducted by an interviewer specialized in the field. Four items were corrected by the researchers to ensure that they were understood by the target users in the sense of the theoretical

model. The group interviews also allowed verification of the ecological validity of the tool (second method) (Brewer, 2000), that is, whether the items and the response choices make sense in the context in which the target users work. Results confirmed that the tool worked properly for partnerships which, beyond providing a space for information sharing and networking, are the venue for collaborative work on a specific project with resources. Finally, the construct validity of the tool (third method) was tested with 28 partnerships totalling 272 respondents. Results were three-fold. First, the tool was able to capture variations in assessments: of the 4,682 responses (272 respondents multiplied by 18 items in the tool), 7.8% indicated conditions not met, 29.1% conditions only partially met, and 63% conditions fully met. Second, the tool allowed for a good convergence between respondents within a same partnership (intra-class correlation coefficient = 0.198, $p < 0.001$). Third, the tool was suitable for distinguishing between partnerships with better and worse results ($F(27; 244) = 3.38$; $p < 0.001$) and helped identify partnerships encountering difficulties with certain conditions.

METHODOLOGICAL GUIDELINES FOR THE TRANSLATION AND VALIDATION OF THE MEASUREMENT TOOLS

The relevant methodological guidelines are drawn from the rules issued by international authorities in this field (ITC, 2017; PISA, 2018; WHO, n.d.). They also come from Quebec scientific literature where expertise has developed because of the widespread use of American measurement instruments requiring translation into French and adaptation to the Quebec context. The methodology established by Vallerand in 1989 constitutes an important reference. These guidelines have been developed mainly in the fields of psychology and education.

Before a measurement tool can be used in a cultural context different from its original context, translation and cultural adaptation are required in order to claim to provide relevant results. This is for linguistic reasons, but also so that the translated tool is as natural and appropriate as the original tool and can work in the same way (ITC, 2017; WHO, n.d.). A verification of the metric properties of the translated tool is also necessary to ensure that it meets quality standards equivalent to the original tool (ITC, 2017). There is thus a distinction between the methodologies for translation and cultural adaptation of the tools, which are detailed here, and those for validation of the translated tools, which we discuss only in terms of their relevance, given the object of the present article.

Translation and Cultural Adaptation of the Measurement Tools

The objective of translation is to ensure equivalence between the translated tool and the original tool. Since it is difficult to match words from one culture to another, given the vocabulary and grammar specific to each language, equivalence of meaning rather than literal translation is targeted (ITC, 2017; PISA, 2018; WHO, n.d.). Caron (1999) distinguishes between equivalence of expressions, retaining terms and expressions specific to the target culture; experiential equivalence,

choosing situations appropriate to the target culture; and conceptual equivalence between the specialized terms of the original tool and the translated tool. Methodologies include translation techniques and pre-testing the draft version with its target users.

Three translation techniques are commonly used: traditional translation, reverse translation, and translation committee. Traditional translation is done by a single bilingual person, the researcher, or a professional translator. This technique alone is not recommended due to the linguistic or comprehension biases inherent to a single translator (ITC, 2017; Vallerand, 1989).

Reverse translation (or back-translation) includes a first translation from the original language into the target language, then a second translation from the translated version into the original language, followed by a comparison of the two versions in the original language, leading to any necessary adjustments. One or two reverse translations can be done in parallel. This technique reduces the risk of bias due to a single translator. The accuracy of the translation in the target language is assessed by the degree to which it accurately reproduces the original version. In this way, gaps in the translation can be identified and corrected. Translators are either professional translators or specialists in the field of the tool who are bilingual and/or familiar with the culture and linguistic characteristics of the target environment (Ouellet, 2008). Adjustments can be identified jointly by the translators (Ouellet, 2008) or by using the committee method described below (Vallerand, 1989). In practice, this method has shown that it is difficult to obtain a perfect equivalence between the translated version and the original version (Caron, 1999). In addition, by not focusing on the target language version of the tool, adaptation problems may be overlooked (Hambleton & Patsula, 1999; ITC, 2017). As such, a parallel double translation with a reconciliation procedure is the preferred approach (Hambleton & Patsula, 1999; ITC, 2017; PISA, 2018).

Translation by a committee involves several bilingual people familiar with the tool domain and the target culture. Committee members may do an initial translation individually or work from one or two original translations (Caron, 1999; Lecavalier & Tassé, 2001). The aim of the committee is to compare the different translations and to establish a consensus version. This technique protects against biases linked to a limited number of translators and makes it possible to verify whether the original version has been adapted to the target culture. It is recognized by many researchers for producing more accurate translations than individual translators (Hambleton & Patsula, 1999). In addition to researchers and translators, it is recommended that the committee include a professional translator or linguist in the target language to ensure proper writing; experts in the field so as to use appropriate specialized terms; and, if possible, the author of the original tool to clarify ambiguities generated by translation (Caron, 1999; Daly, 2010; Vallerand, 1989). The rules for making decisions within the committee are as follows: the meaning of the original item takes precedence over literal translation; the properties of the language as spoken in the target environment must take priority; the accuracy of the technical terms in the target language must

be ensured; in case of disagreement, alternative formulations must be submitted to users for pre-testing; finally, the presentation format and instructions for use of the translated tool must be the same as the original tool (Vallerand, 1989). Because of its advantages, this technique forms a necessary part of the methodology for cross-cultural validation of measurement tools (Vallerand, 1989).

Pre-testing of the draft version (also called the experimental version) is necessary to ensure that the tool items and instructions are clear, unambiguous, and in a language corresponding to that of the target users (Hambleton & Patsula, 1999; ITC, 2017; WHO, n.d.). Two methods are common. The survey method with a sample of target users invites them to point out ambiguous items; those items that are reported repeatedly require reformulation (Lecavalier & Tassé, 2001; Vallerand, 1989). The group interview method with subjects who are representative of the target users also makes it possible to verify whether their understanding of the content corresponds to the intended meaning. This involves asking participants to explain their understanding of each item; if they cannot answer clearly, with the intended meaning, the item is considered ambiguous. In both methods, the pretest should allow participants to clarify the nature of ambiguities and to gather suggestions or opinions on different formulations of certain items. This information is later used by the committee to reformulate problematic items (Caron, 1999; Daly, 2010; ITC, 2017; WHO, n.d.).

Validation of Translated Measurement Tools

The validation of translated tools is used to evaluate their metric properties, compare these properties with those of the original tool, and correct any translation and adaptation problems (Hambleton & Patsula, 1999; ITC, 2017; Vallerand, 1989). Tests may concern the content validity, concurrent validity, construct validity, reliability of the translated tool, or the standards against which to assess its results (Lecavalier & Tassé, 2001; Vallerand, 1989). It is up to the researchers to determine which tests are required based on the nature of the tool and the demonstrated metric properties of the original version. Small-scale studies with the target population are recommended in order to make any necessary revisions to the adapted tool, before proceeding with larger-scale empirical studies on the equivalence of instruments with respect to their metric properties (ITC, 2017).

METHODOLOGY OF TRANSLATION AND CULTURAL ADAPTATION OF THE ORIGINAL FRENCH TOOL

The methodology used in this study is based on the best practices of the above-mentioned methodological guidelines. Two methods were used. First, translation by a committee of experts using parallel double translation was used to translate the tool and verify its equivalence with the original tool. These techniques are favoured by the authorities in the field. Their relevance lies in the group discussion that takes place around the content of the translation, which offers a certain guarantee against the biases of individual translators (Hambleton & Patsula, 1999; ITC, 2017; PISA,

2018). Second, a pretest of the preliminary English version was performed to adapt the tool to the culture of the target users. The group interview method was chosen because it met the main pretest requirement, which was to allow participants to explain the nature of the ambiguities they identify and their suggestions for correcting the tool (Caron, 1999; Daly, 2010; ITC, 2017; WHO, n.d.).

The expert committee members were all bilingual. There were four anglophones—two professional translators, a researcher, and a practitioner (the latter both experts in the field)—and the francophone researcher who authored the original tool. The expert committee had appropriate resources to perform an accurate translation: professional translators who were native speakers of the target language, to ensure the correct linguistic form in English; bilingual experts in the field who were familiar with terminology in both languages, to ensure that the specialized terms used were rigorously reviewed with reference to the theoretical content; and the author of the original tool, to reduce the risk of misunderstandings among committee members and to clarify ambiguities generated by the translation process (Caron, 1999; Daly, 2010; Vallerand, 1989). The role of the expert committee was to evaluate the translations, select the most appropriate wordings, reformulate as necessary, and ensure consistency with the meaning of the original tool. All decisions were made by consensus (always confirmed by vote) among committee members. This procedure led to a preliminary English version of the tool.

Target user adaptation of the tool was carried out by conducting a pretest of the preliminary English version through a focus-group interview with a sample of target users. The focus group was composed of eight participants from Toronto. All had relevant experience in partnership. Three were from public health (two departments) and five were from non-profit groups and community networks (with roles ranging from executive directors to community volunteers). They represented various fields, such as chronic disease prevention, food justice, children's health, women's health, community development, and anti-poverty. The purpose of the pretest was to ensure both that the tool's items were clear, understandable, and unambiguous, in a language corresponding to that of the target users, and that the users' understanding corresponded to the meaning of the original tool. Any item that did not meet these two criteria was reviewed for correction. During the pretest, participants were asked not only to point out ambiguous terms or items but also to provide information (explanations, rewordings) that could correct them. This information was subsequently used by the expert committee to reformulate the problematic items (Caron, 1999; Daly, 2010; Hambleton & Patsula, 1999; ITC, 2017; WHO, n.d.). After a final review, the adapted tool was given final approval by the expert committee.

RESULTS

The translation-adaptation of the tool was conducted from December 2016 to December 2017. In preparation, a proposal for translation-adaptation was developed

Figure 1: The three-step process of translation-adaptation of the original tool

and a partnership agreement was established between Health Nexus of Toronto and the researcher from l'Université de Montréal, who is the author of the original tool. The expert committee for translation was then established. The full process, jointly led by the bilingual expert practitioner (from Health Nexus) and the original tool author (from l'Université de Montréal), was carried out in three steps. The approach for each step is presented below, followed by the results. [Figure 1](#) summarizes the approach for each step.

Step 1: Translation of the Original French Tool and Verification of Its Equivalency by the Expert Committee—June 2017

The tool has three sections: (1) the instructions, which cover the nature of the tool and the six requirements for effective partnership that are evaluated, the three ways to use the tool in a partnership and how to compile the results, and the three types of partnership evaluation that can be produced; (2) collection of information on the respondent (affiliation, role, and time in the partnership); and (3) the 18 items of the tool, with three response options for each.

To begin, two parallel translations of the tool were carried out independently by the committee's two professional translators. This resulted in two very distinct translations, reflecting the differences in the two translators' backgrounds. The first one works full time in translation, but not specifically within the field of the tool; the second one works full time within this field and does internal translation work for Health Nexus. The two translations were the raw material used by the expert committee to develop the preliminary version of the English tool.

The committee edited the translation beginning with the tool's six requirements, then worked through the 18 items, followed by the three options under each item, and finally completed the instructions. This sequence enhanced the accuracy and consistency of the translated texts, from foundation to finish.

The bulk of this work was done in a three-hour face-to-face meeting between all five expert committee members. Both professional translations had been previously submitted to members, asking them to choose their favoured formulations, taking into consideration equivalency in meaning with the original tool, accuracy of English terms commonly used in the field, and clarity of language. At the meeting, a third translation, previously produced by a francophone bilingual professor from the École de santé publique de l'Université de Montréal (who uses the tool professionally), was also used, without the professor participating on the committee. Working together on one screen, the committee proceeded as follows: (1) committee members consulted a three-column table showing the three translations side by side; (2) for each piece (i.e., each requirement or each item), the academic expert first identified a favoured wording to start from; (3) then committee members identified terms and wording to deliberate on and together built a consensus on a preferred formulation; (4) correspondence with the meaning in the French tool was verified and, when needed, the committee went back to reformulate.

The meeting resulted in a consensual choice of texts for the six requirements and 18 items. For five of the 18 items, the committee retained the wording of one of the two professional translations. For the remaining 13 items, as well as the six requirements, the concluding formulation borrowed wording from one or more of the three translations and/or wording proposed by committee members.

The instructions (apart from the six requirements) and the three response options within each of the 18 items were selected and edited from the translated versions over the course of two Skype meetings (three hours each) between two committee members, the practitioner expert, and the author of the original tool. The procedure was the following: (1) while referencing the full committee's chosen formulation for an item, and its three response option translations, the most appropriate terms and wording were identified; (2) correspondence with the meaning in the French tool was then verified; (3) finally, the resulting formulations were validated by email with the remaining expert committee members.

Expert committee members proposed changes for eight of the 54 options (six single options under six separate items and two options under one single item) and four modifications to the instructions. The preliminary English version of the instructions underwent a few notable changes: instructions were spread over two pages instead of one, relying more on point form; and the English term *evaluation* was used to translate the French *diagnostic*, in accordance with the translated title of the tool. The procedure resulted in a consensual formulation of the three response options for each item, and of the instructions. A concluding Skype meeting between the practitioner expert and the author of the original tool was held to integrate all proposed changes and finalize the preliminary English version of the

tool. This English version used the same presentation format and instructions as the original version (except for the two-page length of the instructions), since this condition contributes to the equivalence of the two versions (Vallerand, 1989).

Step 2: Pretest of the Preliminary English Version Tool with a Sample of Target Users—July to October 2017

A two-hour focus group interview was performed with the eight participants. The interview was led by the practitioner expert, with support staff assistance for note-taking. The author of the original tool observed the focus group through remote audio, in order to monitor when and how well participants' understanding corresponded to the meaning of the original tool.

In preparation for the pretest procedure, the practitioner expert asked participants to complete the tool individually by applying it to their partnership. At the beginning of the focus-group session, participants were asked to indicate, by marking on a poster-size version of the tool, any sections that they found difficult to understand. These notes served later as a cue for additional review. Then, for the instructions, as well as for each item and its three response options, the interviewer (1) invited one participant to explain the item in their own words; (2) checked whether other participants had a similar or different comprehension; and (3) for ambiguous statements, comprehension difficulties, or statements that were understood differently from the meaning in the original tool, invited participants to explain the ambiguities and suggest reformulations that improved clarity. Immediately following the focus group, the practitioner expert and the author of the original tool held a Skype meeting to analyze the results. Thereafter, two more Skype meetings were held to produce a final draft of the English translation that incorporated all changes resulting from the pretest. The pretest resulted in the adaptations described below.

Adaptations to the Instructions

Changes were made to the instructions section to clarify how to compile the results. Without changing the presentation of the tool, additional information was presented under a new title: *Value of options*. The instructions section was then resubmitted to the focus group participants by email in October to validate its clarity.

Adaptations to Items and Response Options

Adjustments were made to eight of the 18 items. In items 1, 2, and 3, *issues and options for action* was selected instead of *problem and solution*. Focus-group participants noted that *issue* is a more positive and broader term than *problem* and can reflect the range of situations that mobilize partnerships. Also, these terms better conveyed the original French meaning of *problèmes et solutions*, which refer to the idea of the broad issues that mobilize partners. In items 5, 13, and 18, more precise wording was adopted. The guiding principle for decisions on such changes was that the choice of terms, as understood by the target users, should be informed

Table 1. Adaptations to the tool’s response options and basis for decisions

Response options	Basis of decisions
Item 1, option 2	Standardization
Item 3, option 3	More common terms for target users
Item 4, options 1, 2, 3	Terms/wording closer to the meaning of the original tool
Item 6, option 2	Standardization
Item 8, options 1, 3	Terms/wording closer to the meaning of the original tool
Item 13, options 1, 2, 3	Standardization
Item 14, options 1, 2, 3	More common terms for target users
Item 15, option 3	Terms/wording closer to the meaning of the original tool

by the meaning of the original tool rather than by literal translation (Caron, 1999; Vallerand, 1989). Moreover, in item 2, the term *population* was replaced by *people*. Also, item 14 was reworded to be more direct. For these two items, the guiding principle was to take into account the Canadian English language as spoken by the target users (Vallerand, 1989). There was one instance where the focus group participants identified a translation problem but no acceptable solution. For item 17, they suggested the use of the term *collective impact* to translate *solutions intégrées* and shared a strong consensus on this proposed change. While popular in current discourse on partnership, the term *collective impact* did not accurately translate the meaning, from the original tool, of *solutions intégrées* (which has a broader scope). Instead, the term *holistic* was proposed by the researcher-author and later approved by the full expert committee.

Adaptations were also made to the response options for eight of the items. For five items, these adaptations were made for the same reasons as described above. For three items, the changes were aimed at standardizing terminology used, either between response options or in reference to the item statement. These results are detailed in Table 1.

Step 3: Final Review by the Expert Committee and Publication of the Completed English Tool—November 2017

The resulting final draft of the English tool, with changes tracked since the preliminary English version, was submitted by email to the three other members of the expert committee. These members approved all changes resulting from the pretest, but they added four minor corrections, word changes that refined the adaptation in item 3 (option 2), item 5, item 8 (option 1), and item 15. All committee members approved publication of this final English tool.

DISCUSSION

The methodology used in this study—combining an expert committee process with two parallel translations and a pretest with target users—lends the approach

a high degree of rigour. At each step, this methodology led to improvements in the Canadian English tool. The pretest mainly facilitated changes to more common terms, from the point of view of the target users, and rewordings for greater clarity or a better match with the original tool. The pretest did not result in any major cultural adaptations, reflecting the proximity between the socio-cultural environment of Quebec and English Canada in this field. Through an initial half-day face-to-face meeting of the committee, Skype meetings between the two coordinating committee members, and regular email communications among all members, the expert committee worked effectively to build consensus on the best words and sentence formulations for the three sections of the tool. The rigour of the methodology ensured equivalence of the translated tool with the original tool and solid cultural adaptation to the targeted context (Caron, 1999; Vallerand, 1989; Vézina, Samson-Morasse, Gauthier-Desgagné, Fossard, & Sylvestre, 2011).

When it is empirically demonstrated that a translated version is closely linked to the original version, this supports attribution of the same content validity of the original tool to the translated tool (Vallerand, 1989). Ensuring that items are formulated in equivalent specialized terms, understandable, and culturally appropriate for users is a way to ensure the validity of the information provided by a tool (Ouellet, 2008). Using a committee of experts comprising not only translators but also a researcher and a practitioner familiar with the terminology in both languages, as well as the researcher who produced the original tool, allows us to argue convincingly that the specialized terms used in the English version correspond very closely to the terms and theoretical content of the original tool. During the pretest process, verifying that the target users' comprehension matched the meaning of the items in the original tool was another methodological element that supports the content validity of the translated tool. The construct validity of the translated tool, namely its ability to capture variations in respondents' judgment on the degree to which conditions have been met, to capture expected intra-partnership correlations, and to establish differences between partnerships, will have to be confirmed in a later study.

CONCLUSION

The methodology used was a rigorous process wherein the new version of the tool was adjusted through evaluations by various experts and target users in the field. This ensured that it corresponded to the meaning of the French tool and that it was culturally adapted to its intended users. Of further value, the tool includes indicators that concern a partnership's capacity to direct its multiple stakeholder perspectives towards innovation and includes important process equity indicators that have been raised within many partnership circles in public health. The resulting *Self-evaluation tool for action in partnership* responds well to the evaluation needs of partnerships within English-speaking Canada and other comparable socio-cultural English-speaking regions.

ACKNOWLEDGMENTS

Health Nexus acknowledges the financial support of Status of Women Canada that made this project possible. The authors would like to thank the members who graciously agreed to participate in the expert committee: Robyn Kalda (Health Nexus), Rosalind Lobo (Traductions À la page—translation agency), and Suzanne F. Jackson (University of Toronto, Dalla Lana School of Public Health). They also thank all participants of the focus group: Anonymous (community volunteer), Denise Earle (community volunteer), Elaine Ebach (Toronto Public Health), Julie Maher (Ontario Women's Health Network), Tracey Methven (Toronto Public Health), Talisha Ramsaroop (Social Planning Toronto), Patricia Stevens (Toronto Public Health), and Barbara Willet (Health Nexus). The translation of this article has benefited from the financial support of the REGARDS Team (Fonds de recherche du Québec—Société et culture, 2015-SE-179156) at InterActions—Centre de recherche et de partage des savoirs, CIUSSS du Nord-de-l'Île-de-Montréal. The authors also wish to express their thanks to Patrick Riley for proofreading this article.

NOTE

- 1 Available at <https://en.healthnexus.ca/sites/en.healthnexus.ca/files/resources/selfevaluationtool.pdf>

ORCID ID

Angèle Bilodeau 0000-0002-4400-8624

REFERENCES

- Auspos, P., & Kubisch, A. C. (2004). *Building knowledge about community change: Moving beyond evaluations*. New York, NY: The Aspen Institute. Retrieved from <https://assets.aspeninstitute.org/content/uploads/files/content/docs/rcc/BUILDINGKNOW-ELDGE.pdf>
- Barnes, M., Matka, E., & Sullivan, H. (2003). Evidence, understanding and complexity: Evaluation in non-linear systems. *Evaluation*, 9(3), 265–284. <https://doi.org/10.1177/13563890030093003>
- Bilodeau, A., Galarneau, M., Fournier, M., & Potvin, L. (2011). L'Outil diagnostique de l'action en partenariat: Fondements, élaboration et validation. *Revue canadienne de santé publique*, 102(4), 298–302. Retrieved from <https://link.springer.com/article/10.1007%2FBBF03404054>
- Brewer, M. (2000). Research design and issues of validity. In H. Reis & C. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 3–17). Cambridge, England: Cambridge University Press.
- Caron, J. (1999). *Un guide de validation transculturelle des instruments de mesure en santé mentale*. Rouyn, QC: Université du Québec en Abitibi-Témiscamingue. Retrieved from <http://instrumentspsychometriques.mcgill.ca/instruments/guide.htm>

- Chircop, A., Bassett, R., & Tatlor, E. (2015). Evidence on how to practice intersectoral collaboration for health equity: A scoping review. *Critical Public Health*, 25(2), 178–191. <https://doi.org/10.1080/09581596.2014.887831>
- Dahler-Larsen, P. (2001). From programme theory to constructivism: On tragic, magic and competing programmes. *Evaluation*, 7(3), 331–350. <https://doi.org/10.1177/13563890122209711>
- Daly, Y. (2010). L'adaptation transculturelle d'une échelle de mesure : Cas de l'échelle des ancres de carrière / The transcultural adaptation of a measure scale: Case of the career anchors scale. *Psychologie du travail et des organisations*, 16(2), 160–176. [https://doi.org/10.1016/S1420-2530\(16\)30154-6](https://doi.org/10.1016/S1420-2530(16)30154-6)
- Devaux-Spatarakis, A. (2014). L'évaluation "basée sur la théorie," entre rigueur scientifique et contexte politique. *Politiques & Management Public*, 31(1), 51–68. <https://doi.org/10.3166/pmp.31.51-68>
- Fetterman, D., & Wandersman, A. (2007). Empowerment evaluation: Yesterday, today, and tomorrow. *American Journal of Evaluation*, 28(2), 179–198. <https://doi.org/10.1177/1098214007301350>
- Halliday, J., Asthana, S. N. M., & Richardson, S. (2004). Evaluating partnership: The role of formal assessment tools. *Evaluation*, 10(3), 285–303. <https://doi.org/10.1177/1356389004048279>
- Hambleton, R. K., & Patsula, L. (1999, August). Increasing the validity of adapted tests: Myths to be avoided and guidelines for improving test adaptation practices. *Journal of Applied Testing Technology*, 1–12. Retrieved from <http://www.jattjournal.com/index.php/atp/article/view/48345/39215>
- International Test Commission (ITC). (2017). ITC guidelines for translating and adapting tests (second edition). *International Journal of Testing*, 18(2), 101–134. <https://doi.org/10.1080/15305058.2017.1398166>
- Lecavalier, L., & Tassé, M. J. (2001). Traduction et adaptation transculturelle du Reiss screen for maladaptive behavior. *Revue francophone de la déficience intellectuelle*, 12(1), 31–44. Retrieved from http://www.rfdi.org/wp-content/uploads/2013/05/LECAVALIER_v12.pdf
- Newcomer, K. E., Hatry, H. P., & Wholey, J. S. (2015). *Handbook of practical program evaluation* (4th ed.). Hoboken, NJ: John Wiley & Sons.
- Ouellet, S. (2008). Adaptation et validation d'un instrument de mesure des besoins et des attentes des aidants familiaux en soins palliatifs. *Revue de l'Université de Moncton*, 39(1–2), 97–127. <https://doi.org/10.7202/039844ar>
- Presser, S., Couper, M. P., Lessler, J. T., Martin, E., Martin, J., Rothgeb, J. M., & Singer, E. (2004). Methods for testing and evaluating survey questions. In S. Presser, J. M. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin, & E. Singer (Eds.), *Methods for testing and evaluating survey questionnaires* (pp. 1–22). Hoboken, NJ: John Wiley & Sons.
- Programme for International Student Assessment (PISA). (2018). *PISA 2018 translation and adaptation guidelines*. Prague, Czech Republic: OECD. Retrieved from <https://www.oecd.org/pisa/pisaproducts/PISA-2018-translation-and-adaptation-guidelines.pdf>

- Vallerand, R. J. (1989). Vers une méthodologie de validation trans-culturelle de questionnaires psychologiques : implications pour la recherche en langue française. *Canadian Psychology / Psychologie canadienne*, 30(4), 662–680. <https://doi.org/10.1037/h0079856>
- Vézina, M., Samson-Morasse, C., Gauthier-Desgagné, J., Fossard, M., & Sylvestre, A. (2011). Développement de la version québécoise francophone du Children's Communication Checklist—2 (CCC-2). Traduction, adaptation et équivalence conceptuelle. / Development of a Quebec French version of the Children's Communication Checklist—2 (CCC-2). Translation, adaptation and conceptual equivalence. *Revue canadienne d'orthophonie et d'audiologie* (RCOA), 35(3), 244–253.
- Villeval, M., Bidault, E., Shoveller, J., Alias, F., Basson, J. C., Frasse, C., . . . & Lang, T. (2016). Enabling the transferability of complex interventions: Exploring the combination of an intervention's key functions and implementation. *International Journal of Public Health*, 61(9), 1031–1038. <https://doi.org/10.1007/s00038-016-0809-9>
- World Health Organization (WHO). (2011). *Intersectoral action on health: A path for policy-makers to implement effective and sustainable action on health*. Kobe, Japan: The WHO Center for Health Development. Retrieved from http://www.who.int/kobe_centre/publications/ISA-booklet_WKC-AUG2011.pdf
- World Health Organization (WHO). (n.d.). *Process of translation and adaptation of instruments*. Retrieved from http://www.who.int/substance_abuse/research_tools/translation/en/

AUTHOR INFORMATION

Angèle Bilodeau holds a Ph.D. in applied human sciences and a master's in sociology. She is a full researcher/professor in the School of Public Health at l'Université de Montréal. She is also a researcher at the Centre InterActions, CIUSSS du Nord-de-l'Île-de-Montréal, and a Canada Research Chair in Approches communautaires et inégalités de santé (CACIS). Her research contributions include intersectoral collaborations and governance, action in partnership and social innovation, and effects of intersectoral action on local neighbourhoods.

Gillian Kranias holds a Master in Environmental Studies and a BA in science and human affairs. She specializes in community development and intersectoral partnerships for healthy and equitable communities. She currently works as a bilingual health promotion consultant at Health Nexus, facilitating knowledge mobilization, evaluation, and strategic planning with partnerships.

Comparison of Canadian and American Graduate Evaluation Education Programs

M. Theresa Hunter and James C. McDavid

University of Victoria

Abstract: *University education in evaluation is an important support for the professionalization of the field. Using internet sources and direct contacts with 105 Canadian university departments, our study develops a Canada-wide inventory of graduate courses and programs, with emphasis on evaluation. Comparisons are made with inventories of American graduate evaluation programs undertaken between 1980 and 2018. Although Canada has about half as many multi-course graduate programs in evaluation as were found in the United States in 2018, the relative number of Canadian programs is disproportionately greater, considering the two countries' population sizes. Canada also has a more diverse range of disciplines offering evaluation education programs.*

Keywords: *Canada–US comparisons, evaluation education, inventory, professionalization*

Résumé : *La formation universitaire en évaluation contribue significativement à sa professionnalisation. La présente étude visait à développer un inventaire pan-canadien des programmes et des cours en l'évaluation, et ce, aux cycles supérieurs. La collecte de données a permis de rejoindre 105 départements universitaires en ayant recours aux informations disponibles sur les sites internet ou en contactant des personnes-ressources. Des comparaisons ont été effectuées avec les recensements des programmes d'évaluation américains aux cycles supérieurs entre 1980 et 2018. Bien que le Canada compte environ deux fois moins de programmes aux cycles supérieurs offrant plusieurs cours en évaluation comparativement aux États-Unis en 2018, le nombre relatif de ses programmes est beaucoup plus élevé si nous tenons compte de la taille des deux populations. De plus, le Canada possède également un éventail plus diversifié de disciplines offrant une formation en évaluation de programme.*

Mots clé : *comparaisons Canada – États-Unis, formation en évaluation, inventaire, professionnalisation*

BACKGROUND

The evaluation field has grown significantly since the 1950s and 1960s, when it was first recognized as a distinct field (Worthen, 1990). Supports for the professionalization

Corresponding author: M. Theresa Hunter, School of Public Administration, University of Victoria, PO Box 1700 STN CSC, Victoria, BC V8W 2Y2; thunter@uvic.ca

of evaluation practice have included a growing number of professional associations (LaVelle & Donaldson, 2015), codes of ethics and program evaluation standards (Yarbrough, Shulha, Hopson, & Caruthers, 2011), dedicated journals, listservs (such as the one hosted by the American Evaluation Association [AEA EvalTalk since 1995]), pre-service and in-service training (both credit and non-credit), as well as regional, national, and international conferences. University graduate education in evaluation is one type of support for the professionalization of the field. Inventories of evaluation courses and programs have been undertaken periodically to track the number of evaluation education programs in the United States and internationally (Altschuld, Engle, Cullen, Kim, & Macce, 1994; Conner, Clay, & Hill, 1980; Engle, Altschuld, & Kim, 2006; LaVelle & Donaldson, 2010; May, Fleischer, Scheirer, & Cox, 1986; Stockmann & Meyer, 2016). A preliminary inventory of graduate evaluation courses and programs offered in Canada was conducted by Devine and McDavid in 2009, on behalf of the Consortium of Universities for Evaluation Education (CUEE).

The main purpose of the current study was to identify graduate courses and programs with emphasis on program evaluation offered at universities across Canada. Research questions included the following: how does the Canadian landscape of evaluation education compare with similar inventories of programs in the United States? The full 2018 Inventory of Canadian Graduate Evaluation Education courses and programs is a rich resource for CUEE members, the Canadian Evaluation Society (CES), and others who share an interest in graduate evaluation education. This inventory also provides a baseline for future studies and ongoing monitoring of trends in Canadian evaluation education.

Supports for the professionalization of evaluation in Canada have grown a great deal over the years. Since 1981, CES has grown substantially in its membership and support activities toward the professionalization of evaluation in Canada (CES, n.d.b). CES has held annual conferences since 1980, began publishing *The Canadian Journal of Program Evaluation* in 1986, and has developed and updated professional standards and code of ethics. CES has offered in-service Essential Skills Series since 1999 and continues to update this program and aims to develop more advanced levels of training. CES has offered online webinars since 2012 and created the e-Institute in 2016, to augment online learning opportunities for its members.

A unique feature of the professionalization of evaluation in Canada is the Credentialed Evaluator (CE) designation, established by CES in 2009. The CE designation formally acknowledges professional evaluators' competency in knowledge, skills, experience, commitment, and ongoing learning (CES, n.d.a). A graduate level degree or certificate is a requirement to obtain the CE designation, and curricula that emphasize program evaluation are very desirable (CES, 2010).

In 2008, university partners, in collaboration with CES and the federal government, established the Consortium of Universities for Evaluation Education (CUEE) with the aim of building capacity and expanding opportunities for graduate-level education and training in evaluation across Canada (CUEE, n.d.).

For the past ten years, CUEE has maintained a membership of approximately a dozen universities, representing a diverse range of graduate programs that deliver courses in evaluation in various disciplines, including psychology, education, public administration, and medicine. A growing number of CUEE members offer graduate certificate and diploma credentials in evaluation. These intensive programs help graduates satisfy the requirement for education and training to qualify for the CE designation. While a number of graduate programs offer students the opportunity to specialize in evaluation within their courses of study, opportunities to complete a graduate degree in evaluation are very limited in Canada.

INVENTORIES OF UNIVERSITY-BASED EVALUATION EDUCATION PROGRAMS IN THE UNITED STATES

Periodic studies of university-based evaluation education programs have been undertaken in the United States over the past 40 years (Altschuld et al., 1994; Conner et al., 1980; Engle et al., 2006; LaVelle, 2018; LaVelle & Donaldson, 2010; May et al., 1986). American researchers built upon one another's studies to track the evolution of university-based programs in evaluation at different points in time. Early studies used survey methods to collect data about evaluation courses and education programs, whereas, in later years, the Internet became an important tool in gathering information (e.g., Engle et al., 2006; LaVelle & Donaldson, 2010; LaVelle, 2014, 2018).

American researchers defined a "program" of study in evaluation as "multiple courses, seminars, practicums, offerings, and so on designed to teach evaluation principles and concepts" (Altschuld et al., 1994, p. 72). This definition of "program" was further refined by Engle et al. (2006) and LaVelle and Donaldson (2010) to include university departmental units that offered graduate degree programs with two or more courses with a major emphasis on evaluation. Altschuld et al. (1994) used various methods to construct their initial sampling frame of potential academic programs in evaluation, such as examining lists of programs identified in earlier studies and consulting with representatives of professional associations, including AEA, CES, and the American Sociological Association (ASA). They then conducted a survey to gather detailed information about university graduate evaluation programs in the United States, Canada, and Australia. With an 83% response rate, 49 programs of study were confirmed, including 38 in the United States and 11 in Canada and Australia. These researchers confirmed fewer programs in 1993 in the United States (38) compared with the 44 identified in an earlier study by May et al. (1986), and more in Canada and Australia (11 vs. two in 1986, including nine Canadian universities, one Government of Canada, and one in Australia). Most university evaluation education programs (76%) offered both doctoral and master's degrees, while others (14%) offered a master's degree only. Altschuld et al. (1994) also identified changes in the disciplinary location of evaluation education programs, confirming 67% of programs in the disciplines of education, educational psychology, or psychology, compared

with 85% of programs in these disciplines in 1986. Using the number of courses offered as a proxy for program size, [Altschuld et al. \(1994\)](#) identified 15 (30.6%) small (2–3 courses), 13 (26.5%) medium-sized (4–6 courses), and 21 (42.9%) large (7+ courses) programs.

In a more recent study, [Engle et al. \(2006\)](#) used both a survey and online searches between 2000 and 2002 to gather information about programs of study in evaluation. They sent 84 surveys to candidate programs for which they had contact information (63 US; 23 international), requesting details regarding degrees offered, number of students and their demographics, program FTEs, course descriptions, program goals, areas of specialization, inclusion of evaluation standards, practical experience, and so on. Thirty-eight surveys were returned (45% response rate); however, Engle et al. confirmed just 29 graduate programs in evaluation (24 in the US and five internationally), representing a 41% decline compared with the earlier study ([Altschuld et al., 1994](#)). Among these findings were more small programs (19; 65.5%), fewer medium-sized programs (9; 31.0%), and far fewer large programs (1; 3.4%) than identified in the 1993 survey.

[LaVelle and Donaldson \(2010\)](#) undertook internet searches in 2008 to investigate whether the number of graduate evaluation education programs in the United States was actually in decline. They used a modified definition of “program” to identify university graduate evaluation courses and programs that offered “two or more courses with the word ‘evaluation’ in the course title” (p. 14). The researchers then gathered detailed information regarding a sample of 120 candidate graduate evaluation programs, including the following: disciplines, types of degrees, and the nature of courses. The searches focused on degree programs only and did not include certificate programs: “the search was linked to pages that contain both a degree (MA or MS or PhD) and the term ‘program evaluation’” (p. 13).

Through this search, [LaVelle and Donaldson \(2010\)](#) confirmed 48 programs of study that fit their criteria. Over half (25) of the “sites” offered specialization in evaluation at both the master’s and doctoral levels, while 12 offered specialization only in the master’s degree and 16 in the doctoral degree only. Their findings of 48 programs represented 26% more programs than the 38 identified by [Altschuld et al. \(1994\)](#), and 65.5% more programs than the 29 identified by [Engle et al. \(2006\)](#). [LaVelle and Donaldson \(2010\)](#) determined that 29 programs (60.4%) were located in schools of education, seven (14.5%) in educational psychology, five (10.5%) in psychology, four in public policy (8.3%), and one (2%) in each of criminal justice, applied sociology, and interdisciplinary studies. This finding indicated that 85.4% of programs were located in schools of education, educational psychology, and psychology and was consistent with earlier findings by [May et al. \(1986\)](#), but not [Altschuld et al. \(1994\)](#) or [Engle et al. \(2006\)](#), who respectively found 67% and 62.1% of programs located in these three disciplines. With regard to the size of the 48 programs identified by [LaVelle and Donaldson \(2010\)](#), 31 were small (2–3 courses), 14 were medium-sized (4–6 courses), and three were large (7+ courses) programs. [LaVelle and Donaldson \(2010\)](#) noted that they identified

almost twice as many medium- to large-sized programs and twice as many small-sized programs than [Engle et al. \(2006\)](#). However, the proportion of programs was more in keeping with [Engle et al. \(2006\)](#) than with [Altschuld et al. \(1994\)](#), who found more large programs (42.9% of their sample) and far fewer small programs (30.6% of their sample).

An apparent limitation of [LaVelle and Donaldson's \(2010\)](#) methodology was focusing their online search on courses with “evaluation” in the title, thus eliminating other courses that may have emphasized evaluation in the content but did not contain the word itself in their titles. In previous studies, [Altschuld et al. \(1994\)](#) found that just over half of the 38 programs (not courses) had “evaluation” in the title, while [Engle et al. \(2006\)](#) indicated that 69% of the 29 programs they identified used that word in the program titles.

In his dissertation, [LaVelle \(2014\)](#) used the same internet search methodology as he and Donaldson (2010) had used in their earlier study to identify graduate evaluation education programs in the United States. In addition, he analyzed the extent to which curricula were meeting job demands. [LaVelle \(2014\)](#) identified 87 postsecondary institution “sites” in the United States offering graduate evaluation programs with two or more courses with “evaluation” in their titles, including 55 master’s degrees, 44 doctoral degrees, four specialist degrees, and 36 graduate certificate programs.

[LaVelle's \(2014\)](#) finding of 87 “sites” in the United States offering an estimated 139 evaluation education programs appears to represent 91 more programs in the United States than the 48 identified by [LaVelle & Donaldson in 2010](#) (almost three times as many). In his study, [LaVelle \(2014\)](#) included 36 graduate certificate programs, whereas the 2010 study did not include any certificates. In addition, [LaVelle \(2014\)](#) identified 15 international evaluation education programs, of which six were in Canada, including three certificate programs (offered by the University of Victoria, the University of Ottawa, and Carleton University).

[LaVelle \(2018\)](#) produced an updated directory of evaluation education programs offered in the United States, based on online searches and analysis of publicly available course information undertaken in 2017. The 2018 directory included 71 master’s and doctoral programs and 42 certificate programs. The 113 programs identified is less than the 139 located earlier ([LaVelle, 2014](#)). This directory is not intended to be “definitive,” as it may not include all possible evaluation education programs in the United States. The directory will be corrected as needed and updated every four years. It does not provide analysis regarding the size of programs or their discipline locations.

Survey and internet search methodologies both have strengths and limitations. The survey-only approach requires the predetermination of candidate programs and contacts and relies on respondents to return surveys and provide reliable information. Internet-only searches are limited by the degree to which institutional websites are kept up-to-date and contain accurate information about courses and programs. Using these two methods in combination improves the accuracy of data collection.

2018 INVENTORY OF CANADIAN GRADUATE EVALUATION EDUCATION

The current study used the same definition of “program” of study in evaluation as was used in previous studies ([Altschuld et al., 1994](#); [Engle et al., 2006](#); [LaVelle & Donaldson, 2010](#)) to identify departmental units offering two or more courses, seminars or practica with an emphasis in program evaluation.

Sampling Frame

The sampling frame for this Canadian study was constructed on the basis of several leads, including:

- Current members of the Consortium of Universities for Evaluation Education (CUEE)
- CUEE archival lists of courses and potential member programs
- [Canadian Evaluation Society \(2001\)](#), *Post-Secondary Evaluation-Related Courses*
- [Treasury Board of Canada Secretariat \(2010\)](#), Appendix A—Evaluation Training Program Descriptions
- [Devine and McDavid \(2009\)](#), *CUEE Project: Research on Evaluation Education at the Graduate Level in Canadian Universities, Final Report*
- [Public Health Agency of Canada \(n.d.\)](#), Educational Programs in Public Health
- universitystudy.ca website, using search term “evaluation”
- ReQUE: [Réseau québécois des universitaires en évaluation \(2013\)](#), Inventaire des cours et formations en évaluation de programme
- Internet [Google] search terms: “graduate ‘program evaluation’ course Canada”
- Searches of individual university websites using search terms “evaluation courses,” as well as manual scans of potentially relevant programs
- Intensive, online searches of Canadian university websites between the fall of 2017 and spring of 2018, to identify/confirm graduate evaluation-related courses and programs
- Examination of individual university departmental websites and related graduate course calendars to identify additional courses with emphasis in program evaluation.

The initial sample included 316 candidate courses (both single and multiple course offerings) in 105 academic departments located at 35 universities in nine provinces. Key contacts were identified for each department, including directors, deans, graduate program secretaries, and/or faculty/instructors of evaluation courses.

Data Collection

Initial lists of graduate evaluation courses were compiled for each department and university and then grouped by province and saved as individual pdf documents.

The tentative lists of courses and programs were sent to identified key contacts with a cover email (in either French or English, as applicable) that provided information about the 2018 Evaluation Education Inventory Project and asked them to voluntarily participate in the study. Departmental representatives were asked to verify whether the courses listed were currently being offered and had a major emphasis on program evaluation and to update and make corrections to course numbers, titles, and brief course descriptions. Contacts were also invited to provide information regarding additional graduate courses with a major emphasis on program evaluation offered by their departments.

Response Rate

The overall response rate for 129 departments located at 35 universities was 57% and was similar between predominantly English- and French-speaking universities. Initial lists of courses were revised based on feedback and input received from key contacts. For those courses that were not verified by contacts, additional online searches were conducted to validate that they were currently listed on departmental websites and linked to current and/or future years' graduate course calendars.

Analysis

Penultimate lists of validated evaluation courses were reviewed by two members of the research team (in French and/or English, as applicable) to ensure that all course titles and brief descriptions met the criteria for inclusion in the inventory. These courses covered a wide range of disciplines and subject areas, including economic evaluation and policy development and performance evaluation, with emphasis on program evaluation.

The validated lists of courses were divided into those departments that offered single courses and those that offered two or more courses and therefore qualified as "programs." Departments that offered programs of study in program evaluation were included in the final inventory. Single-course offerings were tracked separately, since they provide an important indication that students are being exposed to "evaluation concepts and processes" (LaVelle & Donaldson, 2010, p. 19).

RESULTS

The complete list of universities, departments, and course titles of the 2018 Canadian Inventory of Graduate Evaluation Education Programs is presented in the Appendix. As shown in Table 1, our study confirmed 239 graduate courses (including some methodology courses) with emphasis in evaluation offered in programs of study (two or more courses) by 54 departments at 27 universities in eight Canadian provinces.

Program Size

In keeping with previous studies (Engle et al., 2006; LaVelle & Donaldson, 2010), the number of courses offered by departments was ranked to determine the size of

Table 1. Number of universities, departments, and courses by province

Province	Universities	Departments	# Courses
British Columbia	3	3	16
Alberta	2	4	11
Saskatchewan	2	4	18
Manitoba	2	3	8
Ontario	8	14	73
Quebec	7	20	88
Nova Scotia	2	2	7
Newfoundland and Labrador	1	4	18
Total	27	54	239

Table 2. Number of courses offered by Canadian university departments

# courses	# departments	Total # of courses
2	13	26
3	12	36
4	10	40
5	5	25
6	4	24
7	3	21
8	3	24
9	3	27
16	1	16
Total	54	239

programs. This proxy measure provides a rough indication of how much emphasis is given to evaluation within graduate programs. [Table 2](#) presents the number of evaluation courses offered by 54 departments at Canadian universities, including 25 small programs (2–3 evaluation courses) (44%), 19 medium-sized programs (4–6 courses) (37%), and 10 large programs (7 or more courses) (19%).

[Table 3](#) identifies the departments, institutions, and locations of the 10 large evaluation education programs, including two in Quebec, five in Ontario, one in British Columbia, one in Saskatchewan, and one in Newfoundland and Labrador.

Range of Disciplines Offering Programs of Study in Evaluation

The current study identified evaluation education programs in 15 different disciplines, as shown in [Table 4](#). The most common types of schools were medicine (9; 17%), education/educational psychology (9; 17%), health-related disciplines (7; 13%), public policy and administration (6; 11%), management studies (4; 7%), social work (4; 7%), and psychology (4; 7%).

Table 3. Details regarding large evaluation education programs in Canada

University	Department	# of Courses
1. Université du Québec, École nationale d'administration publique (ENAP) <i>Québec</i>	École nationale d'administration publique	16
2. University of Toronto <i>Ontario</i>	Institute of Health Policy, Management and Evaluation	9
3. University of Ottawa <i>Ontario</i>	École de gestion Telfer / Telfer School of Management	9
4. University of Waterloo <i>Ontario</i>	School of Public Health and Health Systems	9
5. University of Ottawa <i>Ontario</i>	Faculty of Education, and Faculty of Social Sciences, School of Psychology	8
6. Carleton University <i>Ontario</i>	School of Public Policy and Administration	8
7. Université du Québec à Montréal <i>Québec</i>	École des sciences de la gestion	8
8. University of Victoria <i>British Columbia</i>	School of Public Administration	7
9. University of Regina <i>Saskatchewan</i>	Johnson-Shoyama Graduate School of Public Policy	7
10. Memorial University <i>Newfoundland and Labrador</i>	Faculty of Medicine, Graduate Studies in Medicine	7

Table 4. Academic disciplines offering programs of study in evaluation in Canada

Discipline	No. depts.	%
Medicine	9	17%
Education / Educational psychology	9	17%
Health	7	13%
Public policy / Administration	6	11%
Management	4	7%
Social work	4	7%
Psychology	4	7%
Continuing education	2	4%
Graduate studies	2	4%
Political science	2	4%
Economics	1	2%
Environmental design	1	2%
International affairs	1	2%
Pharmacy	1	2%
Industrial relations	1	2%

The findings of this study can be compared with the results of American research on evaluation education programs. We found only 21% of evaluation education programs in departments of education, educational psychology, and psychology, whereas [May et al. \(1986\)](#) and [LaVelle and Donaldson \(2010\)](#) found 85%, [Altschuld et al. \(1994\)](#) identified 67%, [Engle et al. \(2006\)](#) reported 62.1%, and [LaVelle \(2014\)](#) noted that 79.5% of programs offering specialization in evaluation were located in these three disciplines.

Our study found a greater number and more diverse range of 15 different academic disciplines offering evaluation education programs at Canadian universities compared with studies of programs offered at universities in the United States. Including the top three disciplines offering evaluation education programs (education, educational psychology, and psychology) [LaVelle and Donaldson \(2010\)](#) identified a total of seven different disciplines (with the addition of public policy, criminal justice, applied sociology, and interdisciplinary studies), while [LaVelle \(2014\)](#) found eight (with the addition of public policy, public health, sociology, social work, and business).

Graduate Degrees with Specialization in Evaluation

Our study identified a range of graduate degrees offered by departments with evaluation education courses:

- Most offered degrees at the master's level (e.g., MA, MPA, MPH, MSc)
- 22 departments (41%) offered PhDs;
- Six (11%) offered graduate certificate / diploma programs in evaluation
- Several others offered diploma programs with various specializations (e.g., community health sciences, public policy and administration, public administration)
- Quebec universities frequently offered short, graduate-level microprograms that vary in the number of required courses and credit hours¹

It is not possible to directly compare these findings with studies of evaluation programs in the United States. [Engle et al. \(2006\)](#) found more doctoral programs among the 24 American universities offering specialization in evaluation, with 15 PhDs (62.5%) and 6 EdDs (25%), as well as 15 programs offering MA/MS (62.5%), 6 graduate certificates (25%), and one non-degree program. [LaVelle and Donaldson \(2010\)](#) found over half (52.1%) of the 48 programs offering evaluation education programs were offered for both the master's and doctoral degrees, 12 (18.7%) were offered in master's degree programs only, and 16 (29.2%) were offered in doctoral programs only (this study did not consider certificate programs). Of the 87 post-secondary institutions offering graduate evaluation programs identified by [LaVelle \(2014\)](#), the majority were offered in master's and doctoral degrees, as well as 37 graduate certificate programs. [LaVelle's \(2018\)](#) directory of graduate evaluation education programs included 71 master's and doctoral degrees and 42 certificates

offered by universities in the United States. There appears to be substantial growth in certificate programs in the United States offering evaluation education.

Canadian Graduate Certificate and Diploma Programs in Evaluation

Six Canadian universities offer specialized graduate certificate and diploma programs in evaluation. As shown in [Table 5](#), three of these programs have been in place since the mid-2000s. Three are located in schools of public policy and administration, one is a joint program in education and psychology, one is in management studies, and one is offered by graduate studies. Three of these programs are available online and three provide options to study in French.

Single Course Offerings

In addition to the evaluation programs identified, our study found 61 departments that offer single courses in evaluation within graduate degree programs. As noted, single courses indicate that students are receiving some information about program evaluation concepts and methods ([LaVelle & Donaldson, 2010](#)). The large number of departments offering single courses in evaluation suggests acknowledgment of the relevance of program evaluation across a wide range of disciplines. Some professional associations now have requirements for program evaluation to be included in curricula, such as the [Canadian Psychological Association \(2011\)](#) standards for accreditation of doctoral programs in clinical neuropsychology and clinical, counselling, and school psychology. This indicates a growing recognition of the usefulness of program evaluation in applied settings that may underlie the large number of single courses with emphasis in evaluation across various disciplines.

DISCUSSION

This is the first empirical study to develop a comprehensive inventory of graduate evaluation education programs in Canada. Previous efforts to identify and compile lists of courses and programs in Canada ([CES, 2001](#); [Devine & McDavid, 2009](#); [Treasury Board of Canada Secretariat, 2010](#)) were used to develop the sampling frame for the current study. The [ReQUE \(2013\)](#) Project Report was the most comprehensive listing of evaluation education courses offered by universities in Quebec.

This study identified 54 graduate programs of study offering two or more courses with emphasis in evaluation, with a total of 239 courses at 27 Canadian universities. These programs were offered by 15 different disciplines across all universities. Over half (57%) of the departments that were contacted validated that the courses were currently offered and provided input that ensured they were accurately described. The remaining courses were validated through intensive online searches of university departmental and graduate calendars to confirm they were current and being offered.

Our results indicate that Canada has about half the number of evaluation education programs compared with the most recent findings regarding programs

Table 5. Graduate certificate and diploma credential programs in evaluation

Date of origin	Institution	Department	Delivery	Credential
2006 on-campus 2016 online	Carleton University	School of Public Policy and Administration	Online English	Graduate Diploma in Public Policy and Program Evaluation
2006	University of Ottawa	Faculty of Education, and Faculty of Social Sciences, School of Psychology	On-campus French and English	Graduate Diploma in Program Evaluation / Diplôme d'études supérieures en évaluation et gestion du risque santé des populations
2010	University of Victoria	School of Public Administration	Online English	Graduate Certificate and Diploma in Evaluation
2018	University of the Fraser Valley	School of Graduate Studies	Online English	Graduate Certificate in Program Evaluation
2013	Université du Québec à Montréal (UQÀM)	École des sciences de la gestion [School of Management Sciences]	On-campus French	Diplôme d'études supérieures spécialisées (DEES) en évaluation de programmes, projets et services
pre-2008	Université du Québec	École nationale d'administration publique (ÉNAP)	On-campus, some courses available online French and English	Graduate Diploma in Program Evaluation / Diplôme d'études supérieures spécialisées (DEES), Concentration en évaluation de programmes

located in the United States (LaVelle, 2018); however, Canada has proportionately more programs when the population size of the two countries is taken into account. In 2018, our study identified three times more large graduate programs of study with emphasis in evaluation than were found in the United States in 2008 (10; 19% vs. 3; 6%) (LaVelle & Donaldson, 2010). Canada also has a more diverse range of disciplines offering evaluation education programs, compared with inventories of American evaluation education programs, where the vast majority were in schools of education, educational psychology, and psychology. The most common programs of study in evaluation at Canadian universities in 2018 were located in faculties of medicine, departments of health disciplines, and schools of public policy and administration.

Specialized graduate certificate/diploma programs in evaluation are offered by six different Canadian universities, with three in public policy and administration, one joint program in education and psychology, one in graduate studies, and one in management sciences. Three of these credential programs have existed since the mid-2000s, and three emerged more recently. The number of certificate programs offering evaluation education in the United States appears to be expanding, from six in 2000–2002, initially identified by Engle et al. (2006), to 36 in 2011–2012, noted by LaVelle (2014), and 42 in 2017, reported by LaVelle (2018).

Over the years, the field of evaluation has grown and diversified and become increasingly professionalized. Since 1981, the Canadian Evaluation Society has taken a number of innovative steps to support the professionalization of evaluation in Canada. Requirements to evaluate government programs (Treasury Board, 2016) have increased the demand for qualifications to undertake this work. University partners, such as the CUEE, have sought to build capacity and expand university evaluation education. Graduate evaluation education is now offered by a wide range of departments among Canadian universities, and the number of evaluation courses has expanded, compared to previous listings (Devine & McDavid, 2009; Engle et al., 2006). This indicates a growing awareness of the relevance of evaluation across disciplines.

Constructing inventories is challenging in part because of a tendency for methodologies to yield results that undercount courses, programs, and institutions. By starting with a previously identified population of programs as the basis of the sampling frame, there is the risk of missing programs. Comparisons across countries or over time are subject to this general problem of missing data. Even by using intensive internet searches and following up with surveys of pre-identified programs, there is no assurance that the 2018 Inventory includes all possible programs or evaluation courses in Canada.

The 2018 Canadian Inventory of Graduate Evaluation Education is a resource for CUEE members and other stakeholders with an interest in graduate evaluation education. The inventory will also be of interest to professional associations and governments as they seek to expand the knowledge and skills of evaluators. Finally, the 2018 Inventory will act as a baseline for future studies and ongoing monitoring of graduate evaluation education in Canada.

NOTE

- 1 Examples of microprograms: Laval: 4 to 6 courses (12 to 18 credits); ÉNAP: 3 courses (9 credits).

REFERENCES

- Altschuld, J. W., Engle, M., Cullen, C., Kim, I., & Macce, B. R. (1994). The 1994 directory of evaluation training programs. In J. W. Altschuld & M. Engle (Eds.), *The preparation of professional evaluators: Issues, perspectives, and programs. New Directions for Program Evaluation*, 1994(62), 71–94. <https://doi.org/10.1002/ev.1678>
- Canadian Evaluation Society (CES). (n.d.). *About the CE designation*. Retrieved from <https://evaluationcanada.ca/ce>
- Canadian Evaluation Society (CES). (n.d.). *Our history*. Retrieved from <https://evaluationcanada.ca/our-history>
- Canadian Evaluation Society (CES). (2001, June). *Post-secondary evaluation-related courses*. Unpublished report.
- Canadian Evaluation Society (CES). (2010, April 14). *Professional designations project (archives)*. Retrieved from <https://evaluationcanada.ca/professional-designations-project-archives>
- Canadian Psychological Association (CPA). (2011). *Accreditation standards and procedures for doctoral programmes and internships in professional psychology*, 5th revision. Ottawa, ON: Author.
- Conner, R. F., Clay, T., & Hill, P. (1980). *Directory of evaluation training*. Washington, DC: Pintail Press.
- Consortium of Universities for Evaluation Education (CUEE). (n.d.). *Homepage*. Retrieved from <http://evaluationeducation.ca/>
- Devine, H., & McDavid, J. (2009, June). *Consortium of Universities for Evaluation Education (CUEE) project: Research on evaluation education at the graduate level in Canadian universities*. Unpublished report.
- Engle, M., Altschuld, J. W., & Kim, Y.-C. (2006). 2002 Survey of evaluation preparation programs in universities: An update of the 1992 American Evaluation Association-sponsored study. *American Journal of Evaluation*, 27(3), 353–359. <https://doi.org/10.1177/1098214006288787>
- LaVelle, J. M. (2014). *An analysis of evaluation education programs and evaluator skills across the world* [Doctoral dissertation]. Claremont Graduate University, CA. Retrieved from <http://search.proquest.com/docview/1527108167>
- LaVelle, J. M. (2018). *2018 directory of evaluator education programs in the United States*. Minneapolis, MN: University of Minnesota Libraries Publishing.
- LaVelle, J. M., & Donaldson, S. I. (2010). University-based evaluation training programs in the United States 1980–2008: An empirical examination. *American Journal of Evaluation*, 31(1), 9–23. <https://doi.org/10.1177/1098214009356022>
- LaVelle, J. M., & Donaldson, S. I. (2015). The state of preparing evaluators. In J. W. Altschuld & M. Engle (Eds.), *Accreditation, certification, and credentialing: Relevant*

- concerns for U.S. evaluators. *New Directions for Evaluation*, 2015(145), 39–52. <https://doi.org/10.1002/ev.20110>
- May, R. M., Fleischer, M., Scheirer, C. J., & Cox, G. B. (1986). Directory of evaluation training programs. In B. Gross-Davis (Ed.), *Teaching evaluation across the discipline. New Directions for Program Evaluation*, 1986(29), 71–98. <https://doi.org/10.1002/ev.1421>
- Public Health Agency of Canada. (n.d.). *Educational programs in public health*. Retrieved from http://www.phac-aspc.gc.ca/php-ppsp/master_of_php-eng.php
- Réseau québécois des universitaires en évaluation (ReQUE). (2013, May). *Inventaire des cours et formations en évaluation de programme*. Unpublished report. Retrieved from https://www.sqep.ca/app/uploads/2017/11/Formation_ReQUE.pdf
- Stockmann, R., & Meyer, W. (Eds.). (2016). *The future of evaluation* (pp. 9–19). London, England: Palgrave Macmillan.
- Treasury Board. (2016). *Policy on results*. Ottawa, ON: Government of Canada. Retrieved from <https://tbs-sct.gc.ca/pol/doc-eng.aspx?id=31300>
- Treasury Board of Canada Secretariat. (2010, February 25). *Appendix A—Evaluation training program descriptions*. Ottawa, ON: Centre of Excellence for Evaluation (CEE). Retrieved from <http://www.tbs-sct.gc.ca/cee/tools-outils/qae-aqe/annexa-eng.asp>
- universitystudy.ca. (n.d.). Retrieved from <http://www.universitystudy.ca/search-programs/>
- 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.
- Worthen, B. (1990). Program evaluation. In H. Walberg & G. Haertel (Eds.), *The international encyclopedia of educational evaluation* (pp. 42–47). Toronto, ON: Pergamon Press.

AUTHOR INFORMATION

M. Theresa Hunter, PhD, is an independent consultant in Victoria, British Columbia. Her research interests include social policy regarding children's rights and early childhood education and care, research ethics, and graduate evaluation education. The research for this publication was conducted as a Consortium of Universities for Evaluation Education (CUEE) project.

James C. McDavid is professor emeritus of public administration at the University of Victoria. His research includes topics in program evaluation, performance measurement, and performance management. He has conducted research and program evaluations focusing on federal, state, provincial, and local governments in the United States and Canada. His publications include articles and books on program evaluation (most recently evaluation education), performance measurement, performance management, and comparisons of local government service performance across Canada. The third edition of his textbook, *Program Evaluation and Performance Measurement: An Introduction to Practice*, has recently been published by SAGE Publications in 2019.

APPENDIX

Table A1. 2018 Canadian Inventory of Graduate Evaluation Education

Institution	School / Department	Degrees Offered	Evaluation Course No. / Titles
University of British Columbia	Faculty of Medicine School of Population and Public Health	MSc	SPHA 531 Economic Evaluation
		MHA	SPPH 540 Program Planning and Evaluation
		MHS	SPPH 541 Economic Evaluation
		MPH	SPPH 545 Community Health Promotion
		MPH with Diploma in Dental Public Health	SPHA 553 Program Planning and Evaluation
University of the Fraser Valley	School of Graduate Studies	PhD	
		Graduate Certificate in Program Evaluation	PREV 700 The Context of Policy Development and Performance Evaluation
			PREV 710 Program Evaluation Principles and Processes
			PREV 720 Evaluation Design and Issues in Program Evaluation
			PREV 730 Evaluation Design and Issues in Program Evaluation II
University of Victoria	Faculty of Human and Social Development School of Public Administration	MA	ADMN 537 Program Evaluation and Performance Measurement
		PhD	ADMN 544 Economic Evaluation Methods and Applications
		Graduate Certificate and Diploma in Evaluation	ADMN 580 Qualitative Evaluation Methods and Applications
			ADMN 581 Quantitative Methods for Public Policy Analysis & Program Evaluation
			ADMN 582 Topics in Program Evaluation, Performance Measurement, Performance Management and Knowledge Management
			ADMN 596 Evaluation Project (diploma students)
			ADMN 621 Policy and Program Evaluation and Performance

University of Alberta	School of Public Health	MPH	SPH 504 Health Promotion Planning and Evaluation SPH 631 Health Program Evaluation SPH 695 Epidemiology of Injuries / Design and Evaluation of Injury Interventions
University of Calgary	Werklund School of Education Graduate Programs in Education	MEd MSc	EDER 603.24 Program and Practice Evaluation EDPS 691.04/691.06 Program Development and Evaluation I/ Program Development and Evaluation II
University of Calgary	Cumming School of Medicine Graduate Science Education	MDCH	MDCH 660 Foundations of Health Services Research MDCH 661 (ECON 679) Health Economics I MDCH 662 Economic Evaluation
University of Calgary	Faculty of Social Work	MSW MSW / MBA PhD	MDCH 663 Decision Analysis in Health Economic SOWK 693 Research as a Foundation for Leadership SOWK 695 Becoming an Evidence-Based Leader
University of Regina	Faculty of Education	MEd	EC&I 809 Program Evaluation
University of Regina	Johnson-Shoyama Graduate School of Public Policy	MPA Master's Certificates	EC&I 852 Evaluation in Social Studies Education JSGS 803 Quantitative Methods JSGS 818 Program Evaluation JSGS 824-001 Health Program Planning and Evaluation JSGS 832 Population Based Health Program Management JSGS 838 Public Sector Financial Management JSGS 849 Social Economy and Public Policy JSGS 851 Qualitative Methods
University of Saskatchewan	Department of Psychology Applied Social Psychology	MA PhD	PSY 809.3 Qualitative Research PSY 810 Methods of Applied Social Research PSY 811 Program Evaluation PSY 865.3 Applied Research Designs PSY 902 Practicum in Applied Social Psychology PSY 903 Internship in Applied Social Psychology

(Continued)

Table A1. Continued

Institution	School / Department	Degrees Offered	Evaluation Course No. / Titles
University of Saskatchewan	College of Education Department of Educational Psychology and Special Education	MEd	EPSE 843 Theory of Educational and Psychological Measurement
			EPSE 844.3 Advanced Test Theory and Instrument Construction
			EPSE 993.3 Individual Research Assignment in Educational Psychology
			ECON 7040 Topics in Applied Microeconomics I
University of Manitoba	Department of Economics	Pre-MA MA	ECON 7050 Topics in Applied Microeconomics II
		PhD	ECON 7150 Evaluation of Public Policy and Programs
		Joint MPA	GPOL-7300 Seminar in Theory & Practice of Public Administration: Program Evaluation and the Economic Analysis of Public Programs
University of Manitoba and University of Winnipeg	Department of Political Studies (University of Manitoba) and Department of Political Science (University of Winnipeg)		GPOL-7341 An Indigenous Approach to Program Evaluation: Theory & Practice
			FMILY 7500 Evaluation of Family, Health and Social Development Programs
			CHSC 7290 Economic Evaluation of Health Care
			CHSC 7380 Prevention and Health
University of Manitoba	Rady Faculty of Health Sciences Department of Community Health Sciences	MSc	PADM5420 Policy and Program Evaluation
		MPH	PADM 5441 Introduction to Policy and Program Evaluation
		Diploma	PADM 5442 Quantitative Research Methods in Evaluation
		PhD	PADM 5443 Qualitative Research Methods in Evaluation
			PADM 5444 Benefit-Cost Analysis in Evaluation
			PADM 5445 (Practicum I) Evaluation Research Designs
			PADM 5446 (Practicum II) Data Analysis, Findings and Reporting Methods in Evaluation
			PADM 5715 Policy Research and Evaluation for Indigenous Policy and Administration
Carleton University	School of Public Policy and Administration	Graduate Diploma in Public Policy and Program Evaluation	
		Other Graduate Diploma	
		Specializations	
		MPPA	

Carleton University	Faculty of Public Affairs The Norman Paterson School of International Affairs	MA PhD	INAF 5003 Project Operations in a Developing Context INAF 5609 Development Project Evaluation and Analysis
			EDUC 828 Program Evaluation EDUC-861 Policy Studies in Education EDUC 942 Program Evaluation GDPI-802 (PME 802) Program Inquiry and Evaluation PME-828 Conducting Quality Program Evaluations CSWP 538 Social Work Research: Part I CSWP 638 Social Work Research: Part II CSWP 932 Evaluating Social Work Practices
Queens University	School of Graduate Studies	MEd Graduate Diplomas	
Ryerson University	The G. Raymond Chang School of Continuing Education	Graduate Certificate Course Series in Research Methodologies and Program Evaluation	
University of Guelph	School of Environmental Design and Rural Development (SEDRD)	MSc MPlan	EDRD*6000 Qualitative Analysis in Rural Development RPD*6070 Project Development: Principles, Procedures, and Selected Methods RST*6100 Integrative Research Methods EDRD*6690 Program Evaluation
			CRM 6359 / CRM 6759 Evaluation Of Criminal Justice Programs, Policies and Legislation / Évaluation des programmes, des politiques et des lois en matière de justice criminelle EDU 5299 / EDU 5699 Program Evaluation: Methods and Practice / Évaluation de programmes EDU 5399 / EDU 5799 Development of Assessment and Evaluation Instruments / Elaboration d'instruments d'évaluation
University of Ottawa	Faculty of Education, and School of Psychology, Faculty of Social Sciences	Graduate Diploma in Program Evaluation / DESS en évaluation et gestion du risque santé des populations MEd	

(Continued)

Table A1. Continued

Institution	School / Department	Degrees Offered	Evaluation Course No. / Titles
University of Ottawa	Faculty of Medicine School of Epidemiology and Public Health	MSc Epidemiology	EDU 6191 Methods and interpretation of Quantitative Research
			EDU 6299 / EDU 6699 Program Evaluation: Theory and Contemporary Issues / Évaluation de programmes : théorie et problèmes actuels
			PSY 5104 / PSY 5504 Integration Seminar in Program Evaluation / Séminaire d'intégration en évaluation de programme
			PSY 7102 / PSY 7502 Field Research in Social and Community Interventions / Recherche appliquée aux interventions sociales et communautaires
			PSY 7103 / PSY 7503 Program Evaluation / Évaluation de programmes
			EPI 5144 Global Health Epidemiology and Practice
			EPI 5183 Approaches to Community/Public Health Program Evaluation
			EPI 5189 Health Economic Evaluation
			MBA 5265 Performance Management
			MBA 6156 Performance Management: Business Process Modelling
University of Ottawa	École de gestion Telfer / Telfer School of Management	MHA MBA	MBA 6280 Performance and Operations Management
			MBA 6556 Gestion de la performance : Modélisation des processus de gestion
			MBA 6680 La gestion de la performance et des opérations
			MHA 6203 Program Evaluation for Health Care Managers
			MHA 6215 Management and Evaluation of Quality of Patient Care
			MHA 6301 Population Health and Epidemiology
			MHA 6351 Health Economics

University of Toronto	Institute of Health Policy, Management and Evaluation	MSc PhD	HAD5726H Evaluation and Research Design in Health Informatics
			HAD5728H Performance Measurements in Healthcare: Theory and Application
			HAD5730H Economic Evaluation Methods for Health Services Research
			HAD5738H Advanced Methods for Economic Evaluation
			HAD5740H Intermediate Level Qualitative Research for Health Services and Policy Research
			HAD5742H Mixed Methods for Health Services Research
			HAD5743H Evaluation Design For Complex Interventions
			HAD5780H Program Planning and Evaluation for Health Services and Policy Research
			HAD6761H Health Services Outcomes and Evaluation Comprehensive Course
			CHL5110H Theory and Practice of Program Evaluation
			CHL5150H Data Collection Methods for Research and Evaluation Projects
			CHL7001H Large Scale Impact Evaluation
			EDP3045H Educational Policy and Program Evaluation
			EDP3145H Advanced Issues in Educational Policy Analysis and Program Evaluation
University of Toronto	Faculty of Medicine Department of Physical Therapy	MSc PT	PHT1001H Introduction to Professional PT Practice, Evaluation and Research
			PHT 1006Y Research and Program Evaluation for Physical Therapy Practice I
			PHT 1010Y Research and Program Evaluation for Physical Therapy Practice II
			PHT 1012Y Research and Program Evaluation for Physical Therapy Practice III
University of Toronto	Dalla Lana School of Public Health Social & Behavioural Health Sciences Division Leadership, Higher and Adult Education (LHAE)	MPH PhD	
		MA	
		MEd EdD PhD	

(Continued)

Table A1. Continued

Institution	School / Department	Degrees Offered	Evaluation Course No. / Titles
University of Waterloo	School of Public Health and Health Systems	MSc	PHS 614 Foundations of Program Evaluation
		MPH	PHS 605 Quantitative Methods & Analysis
		MHI	PHS 655 Health Measurement and Survey Methods
		MH Evaluation	PHS 651 Theory and Applications in Program Evaluation
		PhD	PHS 652 Qualitative Methods and Analysis
			PHS 654 Systems Thinking and Analysis in Health Program Planning and Evaluation
University of Windsor	School of Social Work		PHS 653 Evaluation Practice and Management
			PHS 641 Professional Experience Practicum
			HLTH 620 (<i>cross-listed as RLS620 Program Evaluation in Recreation, Sport and Tourism</i>) Experiential Learning in Evaluation
		MSW	47-547 Social Work Research: Practice Evaluation
		MSW / JD	47-625 Law and Social Work: Advanced Practice Research Methods and Policy Analysis
		PhD	47-640 Program Design and Evaluation
Dalhousie University	Faculty of Management School of Public Administration	MPA	47-680 Advanced Internship Seminar
		MPA / JD	MGMT 5006 Program Evaluation, Part 1
		MPA	MGMT 5106 Program Evaluation, Part 2
		MPA/MLIS	MGMT 6525 Program Evaluation
Mount Saint Vincent University	Educational Psychology	MPA/MLIS	MGMT 5146 Research Methods
		Graduate Diploma	PUAD 6520 Program Evaluation Seminar
		MEd	GEPY 6603 Program Evaluation
		GEDU	GEPY 6607 Practicum in Evaluation

Memorial University	Faculty of Medicine Graduate Studies in Medicine	MSc MH Ethics MPH PhD	MED 6100 Teaching and Learning in Medical Education
			MED 6284 Research & Evaluation Design & Methods
			MED 6288 Policy and Decision Making
			MED 6295 Advanced Quantitative Methods
			MED 6950 Simulation and Technology-based Learning in the Health Professions
			MED 6951 Assessment and Evaluation in Health Professions Education
			MED 6953 Current Perspectives and Advances in Medical Education
			ED 6100 Research Designs and Methods in Education
			ED 6205 Educational Policy: Theory and Practice
			ED 6710 Issues in Development and Implementation of Special Education Policy and Practices
Memorial University	Faculty of Education	MED	ED 6822 Foundations of Instructional Design in Post-Secondary Education
			PSY 6400 Theory and Methods in Social Psychology
			PSY 6403 Program Evaluation and Applied Research
			PSY6 404 The Research Project
			PSY 6602 Research Design in Clinical Psychology
			SCWK 6014 Leadership In Social Policy and Programs
			SCWK 6413 Research Theory, Design, and Analysis
			SCWK 6422 Data Analysis for Social Work Assessment and Evaluation
			SCPA 503 Fundamental Skills for CED Practice
			SCPA 511 Field Project—Part II
Memorial University	Faculty of Science Department of Psychology	Masters in Applied Psychological Science PsyD	
Memorial University	School of Social Work	MSW PhD	
Université Concordia	Faculty of Arts and Science, Community Eco- nomic Development (CED) École des affaires pub- liques et communautaires	DESS	

(Continued)

Table A1. Continued

Institution	School / Department	Degrees Offered	Evaluation Course No. / Titles
Université de Montréal	École de santé publique	DESS MSc Microprogrammes	SPU 7501 Évaluation des interventions en santé SPU75031 Practicum en évaluation 1 SPU75032 Practicum en évaluation 2 SPU7504 Séminaire de pratique avancée ASA 6161 Évaluation normative ASA 6175 Méthodes d'évaluation économique ASA 6178 Méthodes d'évaluation ASA 6608 Évaluation économique appliquée à la qualité ASA 6732 Évaluation des technologies de la santé ASA 6814 Évaluation économique
Université de Montréal	École de santé publique Département de gestion, d'évaluation et de politique de santé	Maîtrise en administration des services de santé Maîtrise en santé publique DESS Microprogrammes Doctorat	
Université de Montréal	Département de médecine social et préventive	Maîtrise en santé publique Doctorat en santé publique Diplôme	MSO 6116 Planification et évaluation en santé MSO 6117 Évaluation participative MSO 6136 Introduction à l'évaluation d'impact sur la santé MSO 6331 Introduction à l'évaluation d'impact sur la santé PHA 6320 Ateliers en évaluation de l'information PHM 6014 Évaluation critique de l'essai clinique PSY 6775 Évaluation des programmes d'intervention PSY 7949 Stage: évaluation
Université de Montréal	Faculté de pharmacie Faculté des arts et des sciences Département de psychologie	Maîtrise Doctorat Microprogramme	
Université de Montréal	Faculté des arts et des sciences Département de science politique	Maîtrise	POL 6509 Methodes d'évaluation des politiques publiques POL 6525 Analyses et évaluations de politiques (cours de service MAPI)

Université de Montréal	Faculté des arts et des sciences École de psychoéducation	Maîtrise Microprogramme Doctorat DESS	PSE 6156 Intervention et évaluation 2 PSE 6157 Intervention et évaluation 3 PSE 6575 Méthodologie de la recherche évaluative PSE 6583 Planification et évaluation d'intervention REI 6620 Évaluation de politiques et de programmes en R.I. REI 6625 Audit de la GRH
			ETA 6060 Élaboration et évaluation de programmes ETA 65221 Analyse d'expériences en évaluation 1 ETA 65222 Analyse d'expériences en évaluation 2 ETA 6551 Méthodologie de la mesure en évaluation ETA 7510 Lectures individuelles en mesure et évaluation (crédits 3.0)
			FPE 7802 Introduction à la démarche évaluative FPE 7811 La conduite professionnelle en évaluation de programme ESG 7100 Application des méthodes quantitatives à la gestion FPE 7822 Méthodes qualitatives en évaluation de programme
			SCO 7411 Les indicateurs de performance DSR 7510 Séminaire d'intégration DSR 7521 Stage en évaluation de programme, d'un projet ou d'un service FPE 7842 Approches multiméthodologiques en évaluation de programmes
			ENP 7147 Qualitative Evaluation Methods / Méthodes qualitatives en évaluation de programmes
Université de Montréal	Faculté des arts et des sciences École de relations industrielles	Maîtrise DESS Microprogrammes Doctorat	
Université de Montréal	Faculté des sciences de l'éducation Département de l'administration et fondements de l'éducation	DESS Maîtrise Microprogrammes	
Université du Québec à Montréal	École des sciences de la gestion	DESS Programme court en évaluation de programmes, projets et services	
Université du Québec	École nationale d'administration publique	MPA MSc	

(Continued)

Table A1. Continued

Institution	School / Department	Degrees Offered	Evaluation Course No. / Titles
École nationale d'administration publique (ÉNAP)		Graduate Diploma in Program Evaluation / DEES, Concentration en évaluation de programmes Short program in program evaluation	ENP 7132 + ENP 7132@ Principles, Processes and Evaluation Methods of Public Programs / Principes, processus et méthodes d'évaluation des programmes publics
			ENP 7140 Economic program evaluation methods / Méthodes d'évaluation économique des programmes publics
			ENP 7139 Statistical Methods in Program Evaluation I / Méthodes statistiques d'évaluation de programmes
			ENP 7165 Public-Sector Economics / Economique du secteur public
			ENP 7332 + ENP 7332@ Results-based Management (RBM) / Gestion par résultats
			ENP 7363 Score Board, a Management and Program Monitoring Tool / Tableau de bord, outil de suivi de gestion et de programmes
			ENP 7866 + ENP 7866@ Project Management / Gestion de projets
			ENP 7303 + ENP 7303@ Public Organization Management / Management des organisations publiques
			ENP 7350 Research Methods and Intervention / Méthodes de recherche et d'intervention
			ENP 7505 + ENP 7505@ Principles and Stakes of Public Administration / Principes et enjeux de l'administration publique
			ENP 7521 Public Policy Analysis / Analyse de politique publique
			STA (ENP 7969-S, ENP 7969) ENP 7969-S Work Term/Stage & ENP 7969 Supervised Tutorial / Travail dirigé

Université Laval	Faculté de médecine, Département de médecine sociale et préventive	MSc SAP— évaluation Microprogramme	ENP 7800 + ENP 7800@ Budget Management / Gestion budgétaire
			ENP 7821 Accounting, a Decision-making Tool / Comptabilité, outil de prise de décision
			PIFED Programme international de formation en évaluation du développement
			SAP-7000 Recherche quantitative en santé publique
			SAP-7001 Recherche qualitative en santé publique
			SAP-7013 Évaluation d'impact en santé publique
			SAP-7017 Évaluation économique en santé
			SAP-7006 Évaluation: fondements théoriques et pratiques
			SAP-7020 Mesure et santé: principes théoriques et applications
			SVS-7001 Évaluation de programmes et de l'intervention
Université Laval	École de travail social et de criminologie Faculté des sciences de l'administration Département de management	Maîtrise MBA MSc Maîtrise DESS Microprogramme Doctorat	SVS-7006 Évaluation de l'intervention
			MNG-6003 Recherche évaluative
			MNG-6044 Indicateurs de performance et tableaux de bord de gestion
			MNG-6106 Évaluation : gestion et méthodes
Université Laval	Faculté des sciences de l'éducation Administration et gestion Administration, politiques, mesure et évaluation	DESS Maîtrise Mesure et évaluation Doctorat Mesure et évaluation	AEE-7009 Stage en administration et évaluation en éducation
			ADS-7003 Perspectives critiques sur l'évaluation des politiques éducatives
			MEV-7003 Élaboration / évaluation: programme d'intervention
			MEV-7004 Évaluation de programmes
			MEV-7009 Évaluation institutionnelle

(Continued)

Table A1. Continued

Institution	School / Department	Degrees Offered	Evaluation Course No. / Titles
Université Laval	Faculté des arts et des sciences	Maîtrise	ECN-7180 Évaluation des politiques publiques POL-7013 Évaluation des politiques MEV-7016 Fondements en mesure et en évaluation
Université de Sherbrooke	Administration et gestion Faculté de médecine et des sciences de la santé	Maîtrise en sciences infirmières	RS5741 Évaluation de programmes en santé TXM782 Évaluation d'interventions et de programmes TXM784 Évaluation d'interventions et de programmes SCL708 Évaluation de programme en santé REA113 Évaluation de programme REA110 Démarche évaluative en réadaptation PSE717 Évaluation des programmes PSE836 Conception et évaluation de programmes DFP707 Conception et évaluation de programmes FMED 612 Evaluation Research and Implementation Science (1 credit) PPHS 528 Economic Evaluation of Health Programs PPHS 617 Impact Evaluation
Université de Sherbrooke	Faculté d'éducation	Maîtrise DESS	
McGill University	Faculty of Medicine Epidemiology and Biostatistics	MSc, Family Medicine	

Note. DESS = Diplôme d'études supérieures spécialisées; EdD = Doctorate of Education; GEDU = Graduate Programs in Education; MBA = Master of Business Administration; MDCH = Masters in Community Health Sciences; MEd = Master of Education; MHA = Master of Health Administration; MH Ethics = Master of Health Ethics; MH Evaluation = Master of Health Evaluation; MHI = Master of Health Informatics; MHS = Master of Health Science; MPA = Master of Public Administration; MPA / JD = Master of Public Administration / Juris Doctorate; MPA/MLIS = Master of Public Administration / Master of Library and Information Studies; MPH = Master of Public Health; MPlan = Master of Rural Planning and Development; MPPA = Master of Public Policy and Administration MSc = Master of Science; MSc PT = Master of Science, Physical Therapy; MSc SAP = Maîtrise en santé publique; MSW = Master of Social Work; MSW / JD = Master of Social Work / Juris Doctorate; PhD = Doctor of Philosophy; PsyD = Doctor of Psychology.

Évaluation de la mise en œuvre du programme d'intervention *In vivo* chez les jeunes de 11–12 ans d'une école primaire québécoise présentant un problème de comportements intériorisés : une étude exploratoire

Catherine Fréchette-Simard, Jonathan Bluteau et Isabelle Plante
Université du Québec à Montréal

Résumé : Les troubles intériorisés représentent le problème de santé mentale le plus courant chez les jeunes de 10 à 17 ans. La présente étude exploratoire a évalué, dans une école primaire québécoise, l'implantation du programme *In vivo*, qui s'appuie sur les compétences à gérer le stress pour limiter les comportements intériorisés. Utilisant un devis mixte d'évaluation, les participants au programme ($n = 4$, 11–12 ans), leurs parents ($n = 4$) et des intervenants scolaires impliqués ($n = 7$) ont pris part à l'étude. Les résultats soulignent la pertinence du programme dans ce contexte et révèlent les obstacles et facilitateurs de l'implantation.

Mots clé : anxiété, évaluation de programme, école primaire, prévention, stress, troubles intériorisés

Abstract: Internalized disorders are the most current mental health issue during childhood and adolescence. This study sought to assess the implementation of *In vivo*, an elementary school program aiming to treat anxiety with stress management strategies. The study is based on a mixed method design and collected data from the students experiencing the program ($n=4$, 11-12 y. o.), their parents ($n=4$), and other school staff members involved ($n=7$). The findings of the study show that the program is relevant in this specific context and also highlight the obstacles and facilitators to implementation.

Keywords: anxiety, program evaluation, elementary school, prevention, stress, internalized disorders,

INTRODUCTION

Les problèmes de comportements intériorisés, qui regroupent principalement l'anxiété et la dépression, sont très répandus chez les enfants et adolescents (Le-compte, Moss, Cyr et Pascuzzo, 2014; McLaughlin et Hatzenbuehler, 2009; Narusyte,

Correspondance à l'auteur : Catherine Fréchette-Simard, frechette-simard.catherine@courrier.uqam.ca

Ropponen, Alexanderson et Svedberg, 2017; Zahn-Waxler, Klimes-Dougan et Slatery, 2000). En effet, les troubles anxieux représentent le problème de santé mentale le plus courant chez cette clientèle (Albano, Chorpita et Barlow, 2003; Weissman, Antinoro et Chu, 2009), avec une prévalence pouvant atteindre jusqu'à 33 % des jeunes au Québec et significativement plus élevée chez les filles, tous âges confondus (Piché, Cournoyer, Bergeron, Clément et Smolla, 2017). De plus, les problèmes d'anxiété sont particulièrement comorbides avec les problèmes de dépression : jusqu'à 70 % des jeunes adolescents souffriraient à la fois de symptômes anxieux et dépressifs (Katzman et coll., 2014; Piché et coll., 2017). Lorsqu'ils ne sont pas pris en compte, les comportements intériorisés qui apparaissent durant l'enfance tendent à augmenter et à persister jusqu'à l'âge adulte (Narusyte et coll., 2017), en plus d'entraîner des difficultés relatives à l'estime de soi, aux habiletés sociales (Dumas, 2012) ainsi que des conséquences scolaires négatives (Grills-Taquechel, Fletcher, Vaughn, Denton et Taylor, 2013; Wood, 2006). La période de transition primaire-secondaire représente une période critique pour ces élèves. Non seulement l'arrivée de la puberté constitue-t-elle un moment difficile pour les jeunes, en augmentant leur niveau de stress (Lupien, King, Meany et McEwen, 2001) et leur vulnérabilité à l'anxiété (McLaughlin et Hatzenbuehler, 2009), mais le passage du primaire au secondaire s'accompagne de nombreux changements susceptibles d'accroître la manifestation de comportements intériorisés (ex. changement d'école, perte du cercle d'amis, attentes du milieu scolaire différentes, etc.). Par conséquent, le milieu scolaire, et plus spécifiquement la période qui précède la transition primaire-secondaire, est propice à la prévention des comportements intériorisés chez les jeunes de moins de 12 ans. Dans le cadre de la présente étude, le programme *In vivo* a été implanté pour la première fois dans une école primaire du Québec, auprès de jeunes de 11–12 ans présentant un problème de comportements intériorisés, et la qualité de son implantation a été évaluée. Le choix de ce programme repose notamment sur les résultats prometteurs d'études ayant évalué son implantation auprès d'adolescents de plus de 12 ans (Bluteau, 2017; CISSS, 2016; Dufour et Bluteau, 2016a; Dufour et Bluteau, 2016b; Dufour et Bluteau, 2016c; Girolamo, 2018). Ainsi, évaluer ce programme dans un contexte scolaire primaire avec des élèves plus jeunes est primordial pour valider la faisabilité de celui-ci à l'école primaire et ainsi connaître sa pertinence auprès d'élèves québécois de 11–12 ans présentant une problématique de comportements intériorisés.

CONTEXTE THÉORIQUE

Le programme In vivo

Le programme *In vivo*, développé par le professeur et psychoéducateur Jonathan Bluteau, a vu le jour en 2012 (édité en 2015) à la suite d'un besoin des Centres Jeunesse du Québec d'améliorer les pratiques auprès des jeunes aux prises avec des problèmes intériorisés. En outre, le programme développé devait répondre aux limites des programmes actuels trop centrés sur l'acquisition des connaissances

et peu sur la pratique de nouveaux comportements. Ainsi, le programme vise le développement de compétences pour faire face au stress et s'adresse à des jeunes de 10 à 17 ans présentant des problèmes intériorisés (anxiété, dépression) ou extériorisés (agressivité, impulsivité). Le programme comporte quatre objectifs généraux soit : 1) approfondir les connaissances sur le stress; 2) améliorer l'autorégulation en situation de stress; 3) acquérir des techniques de gestion du stress; et 4) développer l'utilisation des stratégies d'adaptation. Les dix séances du programme, à raison d'une rencontre par semaine, comprennent trois phases : l'éducation au stress, l'éducation et l'entraînement aux stratégies de *coping* (les stratégies d'adaptation) ainsi que l'entraînement par inoculation au stress. Le [tableau 1](#) présente plus en détail les thématiques et le contenu abordés dans le cadre du programme.

Le programme a été qualifié d'innovant par le [ministère du conseil exécutif du Gouvernement du Québec \(2013\)](#), et son originalité est la combinaison de l'approche cognitivo-comportementale avec l'entraînement par inoculation au

Tableau 1. Thématiques et contenu du programme *In vivo*

Thématique par séance	Contenu
1. Brise-glace : première séance activité avec le groupe	Présentation de la structure, des composantes et des objectifs du programme, ainsi que des rôles et responsabilités. Séance-activité réalisée avec les parents
2. À la découverte du stress : la connaissance sur le stress	Le stress, les connaissances sur le stress ainsi que les ingrédients du stress.
3. Je respire par le nez pour ma santé : les impacts négatifs du stress et les techniques de relaxation.	Le stress et la mémoire, l'apprentissage, la santé; la technique de respiration abdominale et la cohérence cardiaque.
4. On va y penser : modèle cognitif et comportemental	L'influence des événements sur les pensées, les émotions et les comportements.
5. Je SPIN mon stress : déconstruction d'une situation stressante	L'analyse d'une situation stressante et la <i>Grille SPIN</i> .
6. Des stratégies pour la vie : les stratégies d'adaptation	Les stratégies positives, négatives et le soutien social pour gérer le stress.
7. Des problèmes aux solutions : la résolution de problème	Les étapes de la résolution de problème et la demande d'aide.
8. Comme les cinq de la main : le soutien social	Le soutien social et le jeu de cartes : <i>Chasse aux mammoths</i> .
9. Mon profil : le profil de l'élève	Les grilles d'échelle de stress cumulées et le profil de stress de l'élève.
10. Mes nouvelles compétences : le bilan	Les réussites et les nouvelles compétences, séance-activité réalisée avec les parents.

stress, lequel consiste à entraîner le jeune en situation de stress réel et contrôlé. En effet, le programme se démarque de la grande majorité des interventions offertes, lesquelles se fondent généralement de façon exclusive sur un enseignement théorique (Barrett et Ollendick, 2004, Fréchette-Simard, Plante et Bluteau, 2018). *In vivo* recourt à l'entraînement par inoculation au stress, à l'aide de modules d'entraînement. Ces modules représentent des activités à prise de risque comme un trampoline, un mur d'escalade, une piscine avec tremplin, etc. Le recours aux modules permet de proposer divers défis aux participants afin de déclencher quatre déterminants du stress : le sentiment de contrôle diminué, la perception de la personnalité menacée, l'imprévisibilité et la nouveauté (Dickerson et Kemeny, 2004; Mason, 1968). La démarche proposée consiste en un véritable entraînement actif à l'autorégulation, ce qui favoriserait davantage le développement des compétences adaptatives (Barlow, Allen et Choate, 2004; Barlow et coll., 2011). Pour pouvoir dispenser le programme, les animateurs doivent, au préalable, avoir suivi une formation théorique et pratique d'une durée de deux jours. De plus, un manuel détaillé, mis à la disposition des animateurs, inclut les objectifs de chaque séance-activité, le matériel nécessaire, le temps dédié à chaque étape du programme, de même que des consignes pour les animateurs, telles que les attitudes et façons de faire à adopter.

À la suite de son implantation, plus de 200 jeunes ont suivi le programme dans divers contextes, notamment dans deux Centres Jeunesse situés au Québec, un Institut thérapeutique éducatif et pédagogique en France et une école secondaire québécoise, le programme *In vivo* a répondu aux critères d'un programme prometteur auprès des adolescents de 12 à 17 ans, présentant un trouble intériorisé (Bluteau, 2017; CISSS, 2016; Dufour et Bluteau, 2016a; Dufour et Bluteau, 2016b; Dufour et Bluteau, 2016c; Girolamo, 2018). Par exemple, Bluteau (2017) rapporte une diminution de l'utilisation des stratégies négatives, de même qu'une diminution significative des symptômes intériorisés chez les garçons. Cependant, le programme n'a pas encore été évalué chez une population d'élèves de moins de 12 ans, ni durant la période de transition entre le primaire et le secondaire. Aussi reste-t-il à évaluer la faisabilité et la qualité de sa mise en œuvre dans un contexte scolaire primaire.

Théorie du programme de Chen

Dans une optique de continuité avec les études évaluatives ayant précédemment évalué le programme *In vivo*, la théorie du programme de Chen (2005) a été ciblée comme cadre de référence pour l'évaluation de la mise en œuvre de ce programme dans la présente étude. Cette théorie est fréquemment utilisée pour l'évaluation de programme puisqu'elle permet de produire de nouvelles connaissances et de mieux comprendre les mécanismes d'action du programme et son processus de changement (Chen, 2005). En effet, ce cadre, schématisé dans la figure 1, permet d'évaluer de façon systématique les différentes variables qui contribueront à l'efficacité d'un programme et de valider la théorie du programme. Chen (2005) évalue, en deux phases distinctes, d'une part la fidélité de l'implantation

Figure 1. Le programme *In vivo* selon le modèle d'action du programme selon Chen

et la faisabilité du modèle d'action du programme et, d'autre part, les effets sur les participants, ce qu'il appelle le modèle de changement. Dans le cadre de la présente étude, seul le modèle d'action a été utilisé, de façon à évaluer chacune des composantes qui contribuent à l'efficacité du programme. Ces composantes sont logiquement liées les unes aux autres et chacune d'entre elles contribue au succès de l'implantation. La figure 1 présente le programme *In vivo* selon le modèle d'action du programme selon [Chen \(2005\)](#), plus précisément il a servi de cadre d'analyse pour la présente évaluation. On y retrouve également le modèle de changement, lequel n'a pas été évalué dans cette étude.

OBJECTIFS DE L'ÉTUDE

Dans un contexte de validation de pratique visant un processus d'amélioration continue du programme, l'objectif général de cette étude consistait à évaluer, dans une visée exploratoire, la mise en œuvre du programme *In vivo* dans une école primaire québécoise, chez des élèves de 6^e année de 11–12 ans présentant une problématique de comportements intériorisés. Plus précisément, l'étude comportait deux objectifs spécifiques : 1) évaluer la fidélité de l'implantation (l'adhésion des animatrices et des participants au programme, le degré d'exposition, la qualité de l'intervention) et 2) évaluer la mise en œuvre du programme à travers les six composantes du modèle d'action de [Chen \(2005, voir figure 1\)](#). Considérant l'absence de données relatives à l'implantation de ce programme au primaire, la présente étude viendra combler un manque évident sur la connaissance du programme et son modèle d'action. De plus, les résultats pourront orienter l'intervention en milieu scolaire primaire afin de limiter les problèmes de comportement intériorisés chez les jeunes de 11–12 ans.

MÉTHODOLOGIE

Afin de répondre aux objectifs fixés, un devis mixte d'évaluation de programme avec échantillon de convenance a été utilisé. L'étude a été réalisée en collaboration avec une commission scolaire et un certificat éthique du comité d'éthique à la recherche de l'Université du Québec à Montréal a été obtenu pour sa réalisation.

Participants

L'échantillon disponible a été limité par certaines contraintes liées au programme ainsi qu'au milieu. En effet, le programme prévoit un ratio de deux animateurs pour quatre participants, alors que le milieu scolaire d'accueil ne disposait pas des ressources pour offrir le programme à plus de quatre participants. L'étude incluait toutefois d'autres acteurs qui ont participé à l'évaluation de la mise en œuvre du programme.

Ainsi, les participants à l'étude, recrutés dans une école de milieu urbain favorisé, étaient 1) quatre élèves (deux filles et deux garçons) de 6^e année du primaire ($M = 11,5$ ans, $ET = 1,0$) qui présentaient un problème de comportements

intériorisés, 2) les parents de ces élèves ($n = 4$, trois femmes et un homme), 3) les animatrices du programme ($n = 2$), 4) les enseignants de ces élèves ($n = 3$, une femme et deux hommes) et 5) la direction de l'école ($n = 2$, deux femmes).

Les participants *élèves* de l'étude ont été recrutés afin de correspondre aux deux critères d'inclusion de l'étude, à savoir l'âge et la présence d'un profil de comportements intériorisés. Ainsi, les élèves ont été sélectionnés parmi une population de 75 élèves de 6^e année, âgés de 11–12 ans. Dix élèves ont d'abord été ciblés par la psychoéducatrice de l'école, en raison de leurs manifestations de comportements intériorisés. Dans le cadre de leur suivi individuel avec leur psychoéducatrice, ces derniers ont été évalués à l'aide des sous-échelles de comportements intériorisés du questionnaire Child Behavior Checklist (CBCL, [Achenbach et Rescorla, 2001](#)). Afin de correspondre au ratio prévu du programme, les quatre participants ayant obtenu le plus haut score au CBCL ont été proposés pour l'étude. Plus spécifiquement, selon les seuils de l'outil, un des élèves retenus atteignait le seuil clinique, deux d'entre eux avaient un score jugé à risque et le dernier avait un seuil sous-clinique. À noter également que l'étude comportait les critères d'exclusion suivants : l'inclusion dans une autre intervention visant le stress ou l'anxiété, la manifestation d'un problème de santé physique ou mentale (autre que l'anxiété ou la dépression) ou un retard cognitif qui aurait limité la participation au programme. Ces critères d'exclusion, déterminés par le programme, sont liés au fait que les activités d'inoculation au stress ont été jugées trop risquées pour des élèves présentant ces profils. Dans la présente étude, aucun élève n'a été exclu sur la base de ces critères.

Les participants *adultes* incluaient les deux animatrices, soit une psychoéducatrice (A1) et une enseignante (A2), qui ont suivi la formation préalable permettant de dispenser le programme. Les autres participants *adultes* (parents, enseignants, direction) ont joué un rôle secondaire dans l'étude, soit dans la documentation de l'appréciation et des effets perçus du programme. Conformément au protocole du programme, un seul parent par élève était présent.

Procédure d'implantation du programme et collecte de données

Le processus de recherche s'est décliné en deux procédures distinctes, soit l'implantation du programme ainsi que la collecte de données. L'intervention a eu lieu sur une période de dix semaines à raison d'une séance de 2 h 30 par semaine entre 14 h 30 à 17 h. Ces séances se sont déroulées chacune en deux phases, la première en salle de classe et la seconde en gymnase dans un centre d'escalade situé à proximité de l'école. Le choix de se tourner vers des activités d'escalade a été fait en raison de l'accessibilité du centre qui se trouvait à distance de marche de l'école. La procédure de collecte de données s'est déroulée en trois phases : 1) avant l'intervention (questionnaires sociodémographique et CBCL), 2) durant l'intervention (questionnaires d'autoévaluation de l'implantation et journal de bord des animatrices, à la fin de chaque séance) et 3) après l'intervention (questionnaires d'autoévaluation de la fidélité et du modèle d'action, questionnaires d'adhésion des participants, CBCL et entrevues). Le détail des différentes mesures utilisées est décrit dans la prochaine section.

Mesures

Un questionnaire sociodémographique a été rempli par les élèves et leur parent avant le début du programme pour recueillir des données concernant le genre du participant, son âge, son niveau scolaire et son état de santé physique et mentale.

Problèmes de comportements intériorisés

Le questionnaire CBCL a été utilisé pour déterminer le profil intériorisé des participants. Pour ce faire, les trois échelles mesurant les comportements intériorisés ($\alpha = 0,91$) ont été retenues, soit l'échelle d'anxiété/dépression, l'échelle de retrait social/dépression et l'échelle de plaintes somatiques. Pour chacun des 32 items de ce questionnaire, l'enfant répond à l'aide d'une échelle de Likert à trois points, allant de « 1 = ne s'applique pas », à « 3 = s'applique toujours ou souvent ». Les participants élèves ont eux-mêmes rempli le questionnaire, sous la supervision de la psychoéducatrice.

Fidélité d'implantation du programme

À la fin de chacune des séances du programme, les animatrices ont répondu à un questionnaire maison d'autoévaluation de l'implantation incluant trois questions ayant pour but de recenser les commentaires concernant 1) les objectifs des séances, 2) la participation des jeunes ainsi que 3) le degré de conformité du contenu par rapport à ce qui est prévu au programme. Dans l'optique de documenter la fidélité de chacune des étapes de l'implantation, les animatrices ont également consigné leurs observations dans un journal de bord.

À la fin du programme, un questionnaire autorapporté évaluant la fidélité de l'implantation (Joly, 2008; Paquette, Joly et Tourigny, 2009) a été rempli par les animatrices. Le questionnaire se divise en deux parties sur la fidélité de l'implantation. La première partie comporte cinq sections : 1) l'adhésion des animateurs, 2) le degré d'exposition (dosage), 3) la qualité de l'intervention, 4) le degré de participation et 5) l'adaptation du programme. La seconde partie se divise en quatre sections, soit le programme, l'organisation, les animatrices ainsi que les participants. Ces deux parties comportent chacun des items auxquels les animatrices doivent répondre à l'aide d'une échelle de Likert en 10 points, allant de « 0 = pas du tout » à « 10 = tout à fait ». Un score de 75 % et plus obtenu aux différents indicateurs d'adhésion, de dosage et de participation correspond à un résultat jugé très fidèle à ce qui était prévu au programme. Un score se situant entre 61 % et 74 % est jugé moyennement fidèle alors qu'un score de 60 % et moins est considéré comme étant peu fidèle (Paquette et coll., 2009). Le questionnaire a procuré des coefficients alpha supérieurs à 0,80 pour toutes les échelles (voir Gamache, Joly et Dionne, 2011).

Dans le but de documenter l'adhésion des participants au programme, les participants élèves ont répondu à une adaptation du questionnaire autorapporté du Child's Evaluation Inventory (CEI) de Kazdin, Siegel et Bass (1992), afin de recueillir de l'information sur leur adhésion au programme et ainsi de connaître la perception de leur expérience. Le score total de l'échelle d'adhésion ($\alpha = 0,89$)

est obtenu à l'aide de deux sous-échelles, soit la Perception des progrès réalisés ($\alpha = 0,95$) et l'Acceptabilité ($\alpha = 0,95$) (Bluteau, 2017). Le questionnaire comporte 19 items fermés, qui comportent des échelles de Likert variées, suivis d'une question à court développement permettant de recueillir des commentaires.

Modèle d'action du programme

Le questionnaire d'autoévaluation du modèle d'action du programme a été utilisé auprès des animatrices (Gamache et coll., 2011). Il comporte six sections documentant : 1) l'implantation auprès d'élèves recevant l'intervention, 2) le processus, 3) les personnes qui implantent le programme, 4) l'organisation, 5) les partenaires ainsi que 6) le contexte écologique. Le questionnaire comporte des questions ouvertes ainsi que des questions à choix de réponse avec des échelles de Likert en 10 points, allant de « 0 = pas du tout » à « 10 = tout à fait ». Ce questionnaire, élaboré et utilisé par Gamache et ses collègues (2011) pour évaluer la fidélité d'implantation d'un programme implanté au Québec, n'a pas encore fait l'objet d'une validation empirique, mais sa validité de construit s'appuie directement sur le choix des questions qui couvrent les six composantes du modèle de Chen (2005).

Des entrevues semi-dirigées ont été réalisées avec les participants *élèves* afin de recueillir des données sur leur expérience. L'entrevue portait sur le programme en général, sur les stratégies d'intervention, sur les changements perçus, sur leur implication personnelle, sur les modules d'escalade ainsi que sur le stress généré par l'escalade.

Les participants *adultes* de l'étude ont répondu à des questions d'appréciation du programme, lors d'entrevues semi-dirigées, afin de documenter leur appréciation de l'intervention ainsi que les effets perçus quant à l'amélioration des comportements des élèves.

Analyses

Pour répondre au premier objectif de l'étude, des scores moyens et écarts-types ont été extraits et calculés à l'aide du logiciel Excel, à partir des données issues des questionnaires de fidélité de l'implantation et des questionnaires d'adhésion des participants. Ensuite, en réponse au second objectif de recherche, chaque composante de l'implantation a fait l'objet d'une évaluation qualitative plus spécifique suivant le modèle de Chen (2005), sous la forme d'une analyse de contenu thématique, à l'aide du logiciel NVivo. Pour ce faire, le contenu des entrevues semi-dirigées a d'abord été retranscrit en verbatim. Ces derniers, de même que les données qualitatives issues des différents questionnaires d'autoévaluation de l'implantation, d'autoévaluation du modèle d'action et du journal de bord ont ensuite été consignés.

RÉSULTATS

Les résultats sont présentés dans les deux prochaines sections, suivant la présentation des objectifs de l'étude, soit l'évaluation 1) de la fidélité de l'implantation et

2) du modèle d'action. Les participants sont ci-après désignés par les abréviations P1, P2, P3 et P4.

Fidélité de l'implantation

Les résultats sont présentés en quatre sections : 1) l'adhésion des animatrices, 2) l'adhésion des participants, 3) le degré d'exposition et 4) la qualité d'intervention. Les scores moyens des différentes sections présentées permettront de qualifier la fidélité de l'implantation en fonction des seuils de [Paquette et coll. \(2009\)](#).

Adhésion des animatrices

Le [tableau 2](#) présente les statistiques descriptives provenant des scores des deux animatrices concernant leur adhésion au programme. Ces résultats montrent, avec un score moyen de 80,8 %, que dans l'ensemble, le cadre théorique du programme a été respecté, de même que les contenus abordés, les activités présentées et l'instrument d'évaluation proposé.

Adhésion des participants

Le score d'adhésion au questionnaire CEI s'est avéré élevé pour P1 (89/95), mais plus bas pour les trois autres participants (P2 = 54/95; P3 = 73/95; P4 = 60/95). En entrevue, tous ont rapporté avoir fait des acquis à la suite du programme et trois des quatre participants ont rapporté avoir préféré l'entraînement en gymnase aux activités structurées. De plus, le score de 6,5/10 pour l'item relatif à l'obtention de la participation attendue révèle, selon les animatrices, que la participation des élèves n'était pas optimale.

Degré d'exposition

Les résultats présentant le degré d'exposition se trouvent dans le [tableau 3](#), qui rapporte un score moyen de 71 %. On constate que la durée des activités a dû être modifiée par rapport à ce qui était prévu (score de 3,5/10). En effet, les données qualitatives ont montré que plusieurs adaptations ont été nécessaires afin de

Tableau 2. Résultats des questions concernant l'adhésion des animatrices

Questions	Moyenne/10 (ÉT)
Les étapes de réalisation des activités ont été suivies.	8,0 (0)
Les principes théoriques en lien avec les activités à réaliser et le changement attendu ont été respectés dans les activités.	9,0 (1,4)
Les contenus spécifiques à aborder ont été respectés.	8,5 (0,7)
Les consignes à suivre lors des activités ont été respectées.	8,0 (1,4)
Les activités prescrites par le programme ont été effectuées.	8,0 (1,4)
Les outils d'évaluations proposés par le programme ont été utilisés.	7,0 (2,8)
TOTAL (sur 60)	48,5 (0,7)
Pourcentage de la moyenne	80,8 %

Tableau 3. Résultats des questions concernant le degré d'exposition au programme

Questions	Moyenne/10 (ÉT)
La durée prévue des activités a été respectée.	3,5 (0,7)
La fréquence prévue des activités a été respectée.	9,0 (1,4)
Les différents types d'activités ont été réalisés.	8,0 (2,3)
La collaboration avec les partenaires a été faite comme prévu.	7,5 (3,5)
L'intervention a été faite comme prévu auprès des différents types de participants (élèves et parents).	7,5 (0,7)
TOTAL (sur 50)	35,5 (2,1)
Pourcentage de la moyenne	71 %

Tableau 4. Résultats des questions concernant la qualité de l'intervention

Questions	Moyenne /10 (ÉT)
Les attitudes particulières identifiées par le programme ont été adoptées.	8,0 (0)
Les manières de faire identifier par le programme ont été respectées.	8,0 (0)
TOTAL (sur 20)	16 (0)
Pourcentage de la moyenne	80 %

s'ajuster aux besoins des participants (ex. : plus de temps pour répondre aux questions posées lors des activités, plus de temps pour s'habiller en hiver) de même qu'aux imprévus relatifs à l'horaire scolaire (ex. : examen à terminer, évènement spécial en classe, etc.).

Qualité de l'intervention

Dans le questionnaire d'évaluation de la fidélité, les animatrices rapportent, avec un score moyen de 80 %, que les attitudes à adopter et les façons de faire prescrites par le programme ont dans l'ensemble été respectées, comme rapporté dans le [tableau 4](#). De plus, en entrevue et dans le journal de bord, les animatrices ont mentionné avoir vécu des difficultés liées à la nouveauté du programme dans leur pratique, à la conciliation des particularités de chaque participant, de même qu'à la gestion des problèmes de discipline.

Modèle d'action

Organisation

Les entretiens réalisés auprès de la direction d'école ont révélé un portrait positif de l'implantation du programme. Toutefois, à la question « La direction a fourni les ressources matérielles suffisantes? » de la section *Organisation* du

Tableau 5. Résultats des questions concernant la dimension Intervenants

Items	Moyenne/10 (ÉT)
J'ai des appréhensions face au programme.	3,5 (4,9)
Je perçois le programme de manière positive.	9,5 (0,7)
Grâce à ce programme, je vais enrichir mon travail.	10,0 (0)
Ce programme va m'apporter des bénéfices personnels.	10,0 (0)
Ce programme aura des effets positifs sur les participants.	8,5 (2,1)
Je crois qu'il est possible d'appliquer un tel programme dans mon milieu.	4,5 (2,1)

questionnaire de fidélité de l'implantation, un score moyen de 6/10 ($\bar{ET} = 5,7$) a été obtenu. Les données du journal de bord des animatrices ont précisé que l'engagement de la direction dans l'implantation a été faible et que peu de supervision a été offerte pour soutenir les intervenantes dans ce processus, notamment parce que le matériel pédagogique nécessaire a été fourni par la présente recherche.

Personnes qui implantent

Le [tableau 5](#) présente les résultats à la dimension *Intervenants* du questionnaire de fidélité de l'implantation. En dépit de scores plutôt élevés, la croyance qu'il est possible d'appliquer un tel programme dans le milieu est plutôt faible, avec une moyenne de 4,5/10 ($\bar{ET} = 2,1$). En entrevue, des difficultés de gestion du temps ont été soulevées par rapport à la tâche planifiée.

Partenaires

La collaboration avec les partenaires a été évaluée à l'aide du questionnaire de fidélité de l'implantation et suggère que celle-ci s'est avérée plutôt positive. Cependant, les animatrices ont toutes deux mentionné, en entrevue et dans le journal de bord, qu'il serait souhaitable de mieux expliquer le programme aux enseignants et de préciser le rôle de chacun des intervenants, afin d'optimiser leur participation en réinvestissant en classe les stratégies apprises.

Contexte écologique

Le journal de bord des animatrices a permis de montrer que dans l'ensemble, le contexte écologique a été favorable à l'implantation du programme, notamment grâce à un local pour la partie théorique ainsi que des collations fournis gratuitement par l'école de même qu'à l'accès gratuit à un centre d'escalade, situé à distance de marche de l'école. Paradoxalement, le fait de devoir quitter l'école a été perçu comme une contrainte importante, étant donné le temps à prévoir pour s'habiller en fonction de la température hivernale. Également, l'environnement du centre d'escalade était peu propice à l'intervention compte tenu de la présence d'autres clients au centre et de l'aménagement physique des lieux : « Nous manquons

d'espace et ce n'est pas l'endroit idéal pour avoir le cartable en main et écrire » (Journal de bord_A2).

Protocole d'intervention

En ce qui a trait à la perception des animatrices quant à la fidélité de l'implantation, les scores ($M = 9,0/10$, $ET = 1,4$) montrent que selon elles, le programme a globalement été implanté comme prévu. Toutefois, les données provenant des questionnaires d'autoévaluation de l'implantation, de même que des entrevues et du journal de bord des animatrices relèvent que plusieurs adaptations ont été faites afin d'adapter le langage à l'âge des participants et d'ajouter du temps supplémentaire pour l'intégration des techniques cognitivo-comportementales. En dépit des adaptations rapportées, les propos des animatrices (entrevues et journal de bord) suggèrent que les interventions prévues dans le protocole du programme étaient adéquates pour la clientèle ciblée : « Je sens toutefois que la partie théorique intéresse bien les élèves et convient quand même bien à leur niveau » (Journal de bord_A2).

Dans un autre ordre d'idées, plusieurs éléments du protocole d'intervention se sont révélés être facilitants, notamment l'activité d'escalade, qui a favorisé la motivation et l'adhésion des participants, ceux-ci ayant nommé l'escalade comme étant l'élément qu'ils préféraient dans le programme. P1 a rapporté en entrevue que les activités de gestion du stress sur les modules d'escalade étaient utiles au transfert des connaissances dans son quotidien. Le journal de bord des animatrices et les questionnaires d'autoévaluation de l'implantation révèlent que l'entraînement a été particulièrement utile pour pratiquer les stratégies d'adaptation enseignées (ex. « travailler fort pour réussir », « demander de l'aide », « avoir une pensée positive », etc.), et moins pour pratiquer les stratégies de gestion du stress (ex. techniques de respiration, de déconstruction du stress, etc.). Contrairement aux propos des animatrices, les entrevues avec les élèves participants et avec leurs parents indiquent plutôt que les stratégies les plus réinvesties par les élèves à la maison sont les techniques de respiration et les stratégies d'adaptation mentionnées précédemment.

Les données du journal de bord soulignent aussi l'impact positif des CFM, par l'aspect attractif pour les jeunes et le soutien à la prise de conscience des réactions de stress sur leur rythme cardiaque. Toutefois, puisque les CFM étaient destinés à des adultes, la lecture de la fréquence cardiaque procurait des résultats instables, nécessitant ainsi plusieurs ajustements pour convenir à la petite cage thoracique des participants.

Clientèle visée

La procédure de sélection, sur la base du questionnaire CBCL, a permis de recruter des jeunes qui présentaient un profil intériorisé, ce que les données du journal de bord ont également permis de corroborer. En effet, P1 présentait principalement un problème au plan de l'anxiété de performance de même qu'un stress quotidien élevé, ce qui avait tendance à augmenter son anxiété. P3 manifestait un

niveau de stress et d'anxiété se traduisant principalement par de la passivité et de l'agitation. Il en était de même pour P4 qui démontrait un niveau de passivité particulièrement élevé durant les séances du programme, en plus de présenter d'autres problématiques non diagnostiquées qui entravaient ses relations avec les adultes et qui se traduisaient par des problèmes de comportement à l'école. P2 manifestait également de l'anxiété et du stress, caractérisés principalement par de l'évitement et un degré de passivité accru face aux difficultés vécues. Les animatrices ont rapporté certains défis liés aux caractéristiques des élèves : « Les enfants choisis ont eu de la difficulté à identifier et verbaliser les situations de stress qu'ils vivaient. » (Questionnaire modèle d'action_A1), « Il a été difficile d'obtenir un niveau d'introspection suffisant de leur part » (Questionnaire modèle d'action_A2). Elles ont également rapporté qu'il y avait beaucoup de discipline à faire et que leur participation aux activités en classe était limitée, principalement à cause de leurs difficultés à s'exprimer. Finalement, la participation des parents, prévue à la première et la dernière séance, a été faible, compte tenu notamment de leur horaire de travail incompatible avec les rencontres prévues et du manque de stationnement à proximité du centre d'escalade.

DISCUSSION

Dans l'ensemble, cette évaluation de programme a permis de documenter la faisabilité du programme *In vivo* pour la clientèle visée, tout en faisant ressortir des pistes d'adaptation à prévoir pour des utilisations ultérieures et la pertinence de poursuivre les études d'implantation dans ce même contexte.

Fidélité de l'implantation

Les résultats montrent que le programme a été implanté de façon relativement fidèle. En effet, les scores obtenus pour l'adhésion des animatrices, et la qualité de l'intervention ont été évalués comme étant très fidèles, avec des scores supérieurs à 75 % (Paquette et coll., 2009). Le score concernant le degré d'exposition (71 %) a été évalué comme moyennement fidèle, puisqu'il se situe entre 61 et 74 %.

Par ailleurs, plusieurs raisons peuvent expliquer que le degré d'exposition ait été jugé moyen, tels le temps perdu durant les périodes de transition, l'importance accrue accordée à la théorie plutôt qu'à la pratique ainsi que la faible participation des élèves, des résultats cohérents avec les études antérieures réalisées avec le programme *In vivo* (Bluteau, 2017; Girolamo, 2018). En dépit des changements apportés quant à la durée destinée aux activités, pratiquement toutes les activités prévues ont été réalisées, soulignant ainsi la faisabilité du programme dans les délais prévus.

Concernant l'adhésion au programme, des écarts importants ont été observés selon les participants, suggérant que les profils initiaux de comportements intériorisés, eux-mêmes plutôt variés, pourraient influencer leur degré d'adhésion. En effet, le participant qui a le plus adhéré au programme semblait présenter des capacités d'introspection et une maturité accrues. Il y aurait donc lieu de vérifier,

auprès d'un plus grand échantillon, si l'âge et la maturité des participants ont une réelle influence sur l'implantation et les effets du programme. Par ailleurs, pour les élèves qui détiennent un profil intériorisé davantage inhibé, certaines recommandations peuvent être émises, notamment de leur donner davantage d'outils pour faciliter leur réflexion face aux sujets abordés.

Organisation, animatrices, partenaires et contexte écologique

Puisque la mise en œuvre du programme s'insérait dans une recherche et que plusieurs ressources matérielles ont été fournies gratuitement, elle n'a pas été réalisée en contexte naturel. Le présent projet ne permet donc pas de confirmer que le milieu possédait les ressources pour mener à terme le programme, ce qui représente pourtant un aspect essentiel d'une implantation réussie (Joly, Tourigny et Thibodeau, 2005). En dépit de ce contexte de mise en œuvre particulier, plusieurs éléments soulevés par Joly et ses collègues (2005) comme étant favorables à une implantation réussie ont été documentés dans la présente étude : l'ouverture et le soutien du milieu scolaire, l'attitude positive et la motivation des animatrices. Toutefois, comme il s'agissait d'une première expérience de mise en œuvre dans cette école et pour cette clientèle, il aurait été bénéfique d'avoir un comité de pilotage afin de soutenir la démarche, comme recommandé dans la formation au programme. Également, de courtes rencontres de suivi ou des communications régulières auraient pu être incluses afin d'informer les enseignants des stratégies vues dans le cadre du programme et ainsi de favoriser le transfert des connaissances dans la classe.

Protocole d'intervention

Les différents résultats présentés précédemment montrent que deux des quatre objectifs généraux du programme semblent avoir été atteints, soit l'objectif d'améliorer les connaissances sur le stress et celui d'améliorer l'utilisation des stratégies d'adaptation. En effet, chacun des quatre jeunes a amélioré ses connaissances sur le stress et l'utilisation de certaines stratégies positives d'adaptation en situation de stress. Malgré le fait que certaines adaptations aient été nécessaires pour bien répondre aux caractéristiques des participants, le protocole semble donc être faisable pour la clientèle ciblée.

Par ailleurs, le protocole d'intervention prévoyait l'utilisation d'une stratégie innovante : l'entraînement par inoculation au stress. Or, la mise en œuvre de cette stratégie a été complexifiée par la difficulté des animatrices à proposer des défis générant suffisamment de stress aux participants lors des activités d'escalade. En effet, selon les formateurs au programme, les expériences d'implantation montrent qu'il faut une expérience d'animation auprès de trois à quatre cohortes avant d'être suffisamment à l'aise avec le matériel (modules et CFM), ainsi que les stratégies d'animation (Bluteau, 2017; CISSS, 2016; Girolamo, 2018). Alors que le programme prévoyait une utilisation accrue des stratégies de gestion du stress (ex. respiration, déconstruction du stress, etc.) sur les modules d'entraînement, ces derniers ont servi davantage à mettre en pratique les stratégies d'adaptation

pour résoudre un problème plutôt que pour gérer le stress. En particulier, la technique de respiration a été difficile à utiliser, possiblement parce que les élèves étaient très intéressés par l'escalade et ainsi moins enclins à prendre des pauses pour réfléchir et mettre en pratique les stratégies enseignées. Ainsi, les mises en œuvre ultérieures devraient prévoir davantage de pauses destinées au processus de réflexion des jeunes, pour ainsi s'assurer d'une acquisition plus optimale des stratégies proposées. De plus, augmenter la complexité des tâches graduellement durant les séances serait souhaitable pour accroître le stress perçu par les jeunes et ainsi favoriser l'inoculation au stress, au cœur du programme *In vivo*.

Les résultats présentés concernant les CFM vont dans le même sens qu'une précédente étude (Bluteau, 2017) qui concluait que les CFM étaient attrayants et utiles. Compte tenu des difficultés d'utilisation rencontrées compte tenu de la petite taille des participants de la présente étude, il serait pertinent de prévoir des CFM mieux adaptés, de grandeur enfant, à ce groupe d'âge.

Clientèle visée

Le fait que la clientèle visée présente des caractéristiques propres à un profil comportemental intériorisé, notamment de la timidité et de l'inhibition, pourrait expliquer certaines des difficultés rencontrées. Ainsi, il est possible que les difficultés vécues par les participants durant le programme aient été liées à leurs caractéristiques personnelles plutôt qu'à leur jeune âge. En ce sens, le programme pourrait potentiellement convenir à une clientèle de 6^e année présentant des comportements intériorisés, moyennant certains ajustements en fonction de leurs capacités d'introspection ou besoins spécifiques.

CONCLUSION

Dans le cadre cette étude exploratoire, le programme *In vivo* a été testé et implanté pour la toute première fois dans un contexte d'école primaire québécoise, auprès d'élèves de 6^e année qui présentent une problématique de comportements intériorisés. L'étude exploratoire évaluant sa mise en œuvre a permis de mieux comprendre les déterminants reliés à la faisabilité d'implantation dans ce contexte. Somme toute, les résultats ont révélé que la fidélité de l'implantation s'est avérée fidèle à ce qui était prévu au programme et que l'évaluation de sa mise en œuvre, à l'aide du modèle de Chen, semble réalisable en contexte primaire, avec la clientèle visée.

En dépit de ces résultats positifs, la présente étude comporte certaines limites qu'il convient de reconnaître et qui pourront orienter la recherche future. Une première limite de cette évaluation concerne la petite taille d'échantillon, qui découle principalement des contraintes du programme *In vivo*, destiné à quatre participants. D'autres évaluations incluant de plus grands échantillons seront donc nécessaires, pour évaluer la généralisabilité des résultats auprès de clientèles diversifiées pour ainsi confirmer la validité du programme *In vivo* en contexte primaire. De plus, certaines adaptations seraient souhaitables afin de favoriser l'atteinte des objectifs ciblés par le programme, tels des ajustements pour mieux tenir

compte des besoins des élèves inhibés, une meilleure organisation des contenus théoriques et pratiques et des mesures additionnelles pour favoriser l'implication des partenaires. Une autre limite de l'étude provient du fait que l'évaluation du programme repose notamment sur une autoévaluation rapportée par les animatrices du programme. Ainsi, dans des recherches futures, il serait pertinent de trianguler ces données avec celles provenant d'un observateur externe, exempt de biais. En dépit de ces limites, la mise en œuvre de ce programme souligne la faisabilité des interventions de groupe en milieu scolaire primaire, une avenue de choix pour limiter le développement des problèmes liés à l'anxiété à l'école. Sur la base des recommandations qui découlent de cette première implantation, les milieux scolaires primaires pourront donc adapter celui-ci à leur convenance et selon leurs besoins, pour ainsi contribuer à limiter les problématiques liées aux comportements intériorisés.

BIBLIOGRAPHIE

- Achenbach, T. M. et Rescorla, L. A. (2001). *Manual for the ASEBA school-age forms & profiles: An integrated system of multi-informant assessment*. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.
- Albano, A. M., Chorpita, B. F. et Barlow, D. H. (2003). Childhood anxiety disorders. Dans E. J. Mash et R. A. Barkley (éds.), *Child psychopathology* (pp. 279–329). New York, NY: Guilford Press.
- Barlow, D. H., Allen, L. B. et Choate, M. L. (2004). Toward a unified treatment for emotional disorders. *Behavior Therapy*, 35(2), 205–230. [https://doi.org/10.1016/s0005-7894\(04\)80036-4](https://doi.org/10.1016/s0005-7894(04)80036-4)
- Barlow, D. H., Farchione, T. J., Fairholme, C. P., Ellard, K. K., Boisseau, C. L., Allen, L. B. et Ehrenreich-May, J. (2011). *Unified protocol for transdiagnostic treatment of emotional disorders: Therapist guide*. New York, NY: Oxford University Press.
- Barrett, P. et Ollendick, T. H. (2004). *Handbook of interventions that work with children and adolescents: Prevention and treatment*. Chichester: John Wiley et Sons Ltd.
- Bluteau, J. (2017). *Validation d'un programme de développement de compétences pour faire face au stress chez des adolescents présentant des troubles intériorisés*. Thèse de doctorat. Université du Québec à Trois-Rivières.
- Chen, H. (2005). *Practical program evaluation: Assessing and improving planning, implementation and effectiveness*. Thousand Oaks, CA: Sage Publications.
- CISSS Centres jeunesse de Laval. (2016). *Programme In vivo : évaluation de la mise en œuvre*. Laval, Québec: CISSS Laval.
- Dickerson, S. S. et Kemeny, M. E. (2004). Acute stressors and cortisol responses: A theoretical integration and synthesis of laboratory research. *American Psychological Association*, 130(3), 355–391. <https://doi.org/10.1037/0033-2909.130.3.355>. Medline:15122924
- Dufour, M.-M. et Bluteau, J. (2016a). *Évaluation de l'implantation du programme In vivo Institut Thérapeutique, Éducatif et Pédagogique d'Angoulême*. Montréal, Québec: UQAM.
- Dufour, M.-M. et Bluteau, J. (2016b). *Évaluation de l'implantation du programme In vivo : CISSS-Centres jeunesse de Lanaudière*. Montréal, Québec: UQAM.

- Dufour, M.-M. et Bluteau, J. (2016c). *Évaluation de l'implantation du programme In vivo : CISSS-Centres jeunesse des Laurentides*. Montréal, Québec: UQAM.
- Dumas, J. E. (2012). *L'enfant anxieux. Comprendre la peur de la peur et redonner courage*. Bruxelles: De Boeck Éditions.
- Fréchette-Simard, C., Plante, I. et Bluteau, J. (2018). Strategies included in cognitive and behavioral therapy programs to treat internalized disorders: A systematic review. *Cognitive Behaviour Therapy*, 47(4), 262–285. <https://doi.org/10.1080/16506073.2017.1388275>. Medline:29103355
- Gamache, V., Joly, J. et Dionne, C. (2011). La fidélité d'implantation du programme québécois d'intervention comportementale intensive destiné aux enfants ayant un trouble envahissant du développement en CRDITED. *Revue de psychoéducation*, 40(1), 1–23.
- Girolamo, A. (2018). *Une meilleure capacité à faire face au stress pour une réduction de l'agressivité réactive chez les élèves du secondaire : évaluation du programme In vivo*. Mémoire de maîtrise. Université du Québec à Montréal.
- Grills-Taquechel, A. E., Fletcher, J. M., Vaughn, S. R., Denton, C. A. et Taylor, P. (2013). Anxiety and inattention as predictors of achievement in early elementary school children. *Anxiety Stress Coping*, 26(4), 391–410. <https://doi.org/10.1080/10615806.2012.691969>. Medline:22764776
- Joly, J. (2008). *Questionnaire d'autoévaluation de la fidélité d'implantation d'un programme*. Communication personnelle, Mai, 2008.
- Joly, J., Tourigny, M. et Thibaudeau, M. (2005). La fidélité d'implantation des programmes de prévention ou d'intervention dans les écoles auprès des élèves en difficulté de comportements. *Nouveaux cahiers de la recherche en éducation*, 8(2). 101–110. <https://doi.org/10.7202/1017533ar>
- Katzman, M., Bleau, P., Blier, P., Chokka, P., Kjernisted, K., Van Ameringen, M., ... l'Université McGill. (2014). Canadian clinical practice guidelines for the management of anxiety, posttraumatic stress and obsessive-compulsive disorder. *BioMed Central Psychiatric*, 14(Suppl 1), 1–83. <https://doi.org/10.1186/1471-244x-14-s1-s1>. Medline:25081580
- Kazdin, A. E., Siegel, T. C. et Bass, D. (1992). Cognitive problem-solving skills training and parent management training in the treatment of antisocial behavior in children. *Journal of Consulting and Clinical Psychology*, 60(5), 733–747. <https://doi.org/10.1037//0022-006x.60.5.733>. Medline:1401389
- Lecompte, V., Moss, E., Cyr, C. et Pascuzzo, K. (2014). Preschool attachment, self-esteem and the development of preadolescent anxiety and depressive symptoms. *Attachment and Human Development*, 16(3), 242–260. <https://doi.org/10.1080/14616734.2013.873816>. Medline:24417266
- Lupien, S. J., King, S., Meany, M. J. et McEwen, B. S. (2001). Can poverty get under your skin? Basal cortisol levels and cognitive function in children from low and high socioeconomic status. *Development and Psychopathology*, 13(3), 653–676. <https://doi.org/10.1017/s0954579401003133>. Medline:11523853
- Mason, J. W. (1968). A review of psychoendocrine research on the pituitary-adrenal cortical system. *Psychosomatic Medecin*, 30(5), 576–607. <https://doi.org/10.1097/00006842-196809000-00020>. Medline:4303377

- McLaughlin, K. A. et Hatzenbuehler, M. L. (2009). Stressful life events, anxiety sensitivity, and internalizing symptoms in adolescents. *Journal of Abnormal Psychology*, 118(3), 659–669. <https://doi.org/10.1037/a0016499>. Medline:19685962
- Ministère du conseil exécutif du Gouvernement du Québec. (2013). Le programme *In vivo* : une approche innovante en santé mentale chez les jeunes. *Bulletin prospective*, 17(1). Récupéré le 9 décembre 2015 de. <https://w3.mce.gouv.qc.ca/SPPS-Bulletin/AfficherArticle.asp?Cle=80>
- Narusyte, J., Ropponen, A., Alexanderson, K. et Svedberg, P. (2017). Internalizing and externalizing problems in childhood and adolescence as predictors of work incapacity in young adulthood. *Social Psychiatry and Psychiatric Epidemiology*, 52(9), 1159–1168. <https://doi.org/10.1007/s00127-017-1409-6>. Medline:28638965
- Paquette, G., Joly, J. et Tourigny, M. (2009). La fidélité de l'implantation d'un programme d'intervention pour adolescentes agressées sexuellement : une mesure par observation systématique. *The Canadian Journal of Program Evaluation*, 24(2), 57–79.
- Piché, G., Cournoyer, M., Bergeron, L., Clément, M. et Smolla, N. (2017). Épidémiologie des troubles dépressifs et anxieux chez les enfants et les adolescents québécois. *Santé mentale au Québec*, 42(1), 19–42. <https://doi.org/10.7202/1040242ar>
- Weissman, A., Antinoro, D. et Chu, B. (2009). Cognitive-behavioral therapy for anxious youth in school settings. Advances and challenges. Dans M. Mayer, R. Van Acker, J. Lochman, et F. Gresham (éds.), *Cognitive-behavioral interventions for emotional and behavioral disorders: School Based Practice* (pp. 173–203). New York, NY: Guilford Press.
- Wood, J. (2006). Effect of anxiety reduction on children's school performance and social adjustment. *Developmental Psychology*, 42(2), 345–349. <https://doi.org/10.1037/0012-1649.42.2.345>. Medline:16569172
- Zahn-Waxler, C., Klimes-Dougan, B. et Slattery, M. J. (2000). Internalizing problems of childhood and adolescence: Prospects, pitfalls, and progress in understanding the development of anxiety and depression. *Development and Psychopathology*, 12(3), 443–466. <https://doi.org/10.1017/s0954579400003102>. Medline:11014747

PRÉSENTATION DES AUTEURS

Catherine Fréchette-Simard complète actuellement un doctorat en éducation à l'Université du Québec à Montréal. Ses recherches portent sur les comportements intériorisés et l'anxiété de performance à l'école primaire et secondaire, de même que les liens entre ces problématiques et la motivation scolaire de l'élève. Elle est également coordonnatrice de la Chaire de recherche sur l'égalité des genres à l'école et œuvre à différents projets portant sur cette thématique.

Jonathan Bluteau est psychoéducateur et professeur agrégé au département d'éducation et formation spécialisées à l'Université du Québec à Montréal (UQAM). Le professeur Bluteau détient un doctorat en psychoéducation et se spécialise dans l'élaboration et l'évaluation de programmes d'intervention pour adolescents ayant des troubles anxio-dépressifs. Ses intérêts de recherche ciblent la prévention des problèmes de santé mentale chez les jeunes, l'étude des biomarqueurs du stress liés à la santé mentale, ainsi que le phénomène du surdiagnostic et du surtraitement médical en santé mentale. Sa posture, ses projets

d'intervention et ses projets de recherche s'appuient sur les théories de l'évolution et la biologie évolutive. Récemment, il s'intéresse aux impacts de l'architecture des écoles et l'aménagement de l'environnement de la classe sur le bien-être des élèves et des enseignants. Il travaille et intervient auprès de trois réseaux en France, soit l'Association française de thérapie cognitive et comportementale, le réseau des Instituts thérapeutiques éducatifs et pédagogiques, ainsi que l'Association française des psychologues de l'éducation nationale. Il est membre de l'Ordre des psychoéducateurs et psychoéducatrices du Québec, de l'Institut santé et société de l'UQAM, de la Société québécoise d'évaluation de programme (SQEP) et de la Société canadienne d'évaluation (SCE). Également, il est chercheur régulier au Centre d'études sur l'apprentissage et la performance UQAM (CEAP)

Isabelle Plante est professeure au département d'éducation et formation spécialisées de l'Université du Québec à Montréal et chercheure régulière au Centre d'études sur l'apprentissage et la performance (CEAP). Titulaire de la Chaire de recherche sur l'égalité des genres à l'école (CRÉGÉ), ses travaux se centrent sur la compréhension des facteurs qui caractérisent les parcours scolaires différenciés des garçons et des filles du primaire et du secondaire. Elle s'intéresse notamment au rôle des stéréotypes sociaux de genre à propos des capacités cognitives selon le sexe dans la motivation, la réussite et les aspirations professionnelles des garçons et des filles. Ses recherches traitent également des problématiques typiquement vécues selon le genre à l'école, comme l'anxiété de performance chez les filles et les troubles de comportement extériorisés chez les garçons.

Dramatizing Learning, Performing Ourselves: Stories of Theatre-Based Evaluation in Vancouver's Downtown Eastside

Christopher Cook, Carolyn Camman, Andrew Leyland, Suzie O'Shea, and
Angela Towle

UBC Learning Exchange, University of British Columbia

Abstract: *Voices UP!*, a play developed and performed collectively with participants from Vancouver's Downtown Eastside, illustrates the use of theatre as an innovative evaluation method for data collection, analysis, and knowledge translation. This artful process can convey complex, experiential evaluation findings and create engaging opportunities for learning, while building relationships and skills among participants. In this article we describe the creation process utilized for this theatre-based evaluation project, as well as guiding principles and lessons learned for evaluators who may want to engage in similar theatre-based participatory work.

Keywords: arts-based evaluation, knowledge translation, participatory evaluation, research-based theatre

Résumé : *Voices UP!*, une pièce de théâtre communautaire créée et interprétée collectivement avec des participants du Downtown Eastside de Vancouver, illustre l'utilisation du théâtre comme méthode d'évaluation novatrice pour la collecte de données, l'analyse et l'application des connaissances. Ce processus artistique peut transmettre des résultats d'évaluations complexes et expérientiels et créer des mécanismes d'apprentissage intéressants, tout en construisant des relations et des compétences entre les participants. Dans cet article, nous décrivons le processus de création employé pour ce projet d'évaluation fondé sur le théâtre, ainsi que les principes directeurs et les leçons apprises pour les évaluateurs et évaluatrices qui souhaiteraient s'engager dans un travail participatif similaire basé sur le théâtre.

Mots clé : évaluation fondée sur les arts, traduction des connaissances, évaluation participative, recherche basée sur le théâtre

SETTING THE SCENE

In 1999, the University of British Columbia (UBC) founded the Learning Exchange to foster engagement between the university and communities of Vancouver's

Corresponding Author: Angela Towle, Academic Director, UBC Learning Exchange, 612 Main Street, Vancouver, BC V6A 2V3; angela.towle@ubc.ca

Downtown Eastside (DTES) (Towle & Leahy, 2016). The DTES is an inner-city area made up of a diverse set of overlapping neighbourhoods, rich in history and art, many of whose residents face multiple barriers—including homelessness, poverty, and gentrification, as well as negative public perceptions and stigma (City of Vancouver, 2014; Eagle Bear, 2016). To serve the needs of the DTES community, the Learning Exchange brings together local residents, university students, and faculty members with a range of programming rooted in asset-based community development (Towle & Leahy, 2016), knowledge sharing, and skill building—programs such as the Learning Lab.

Experimental from the beginning, the Learning Lab was created in 2013 as a low-barrier “middle space” between more formal learning opportunities, such as structured English conversation classes, and a drop-in style environment. Its aim is to motivate community members to try new things, build relationships, and explore leadership roles. Ideas for Learning Lab activities were gathered directly from the community, and took the form of ongoing clubs—such as chess, Mahjong, and math clubs—as well as diverse workshops, from robotics to button-blanket making. Reflecting a strengths-based approach, the Learning Lab aimed to create activities that could be self-sustaining, run by and for community members who were engaged as fellow learners and facilitators, with the support of Learning Exchange staff and university students. Learning Lab staff define leadership on a continuum, acknowledging that leading may look different for each community member, and encouraging learners and facilitators to discover what leadership means to them and to lead from where they are at.

We undertook an evaluation of the Learning Lab in 2015 to determine program outcomes and future directions. The objectives that required evaluation included to what extent the program (a) established a range of “middle space” educational activities that supported weekly participation by community members; (b) built new connections with external partners that supported diverse programming; and (c) developed activities cooperatively sustained and facilitated by local residents (Leyland, 2016). In accordance with the program's values, we sought evaluation methods that were participatory, creative, and fun and would support community engagement. In addition to a larger program evaluation that drew on a variety of more traditional quantitative and qualitative methods and a mixture of data sources, from semi-structured interviews to attendance sheets (Leyland, 2016), we initiated a theatre-based evaluation project. The Learning Exchange had previously explored ethnotheatre as a means of capturing program participants' voices (Balyasnikova, Higgins, & Hume, 2017; Gillard & Balyasnikova, 2015), and the Learning Lab had facilitated several collaborative art-making projects. Drawing on these experiences, and rooting the work in participatory evaluation approaches (Chouinard, 2013; Cousins & Earl, 1992; Cousins & Whitmore, 2007), we began creating a play called *Voices UP!*. Community members wrote and performed themselves in this play, side-by-side with staff, including the program evaluator, who co-created and appeared in the production. We selected theatre-based evaluation as an approach to allow community members to participate in a

responsive, engaging evaluative process that centred on learning (Patton, 1994), while providing an accessible medium through which the evaluation narratives could be shared. Although the theatre-based project began as a component of the larger program evaluation of the Learning Lab (Leyland, 2016), it became an independent evaluation journey that lasted well-beyond the completion of the program evaluation (see Figure 1).

We describe the creation process utilized in the Learning Lab's theatre-based evaluation project, as well as guiding principles and "lessons learned" for evaluators who may want to engage in similar participatory work. In sharing our process, we offer a distinctive, creative map for engaging in collaborative arts-based evaluation with community members. A glossary of theatre terms is provided in the Appendix.

WHY THEATRE? APPLICATIONS OF THEATRE IN EVALUATION AND RESEARCH

There have been recent calls for the increased use of creative methods in evaluation (Searle & Shulha, 2016; Simons & McCormack, 2007), but—with the exception of Cozart, Gordon, Gunzenhauser, McKinney, and Petterson (2003) and Barone (1991)—the literature on the use of theatre in evaluation remains sparse. In other disciplines, research-based theatre is a methodology supported by substantial scholarship (Belliveau & Lea, 2016), for example, in education (Beck, Belliveau, Lea, & Wager, 2011; Belliveau, 2007; Donmoyer & Yennie-Donmoyer, 1995; Norris, 2000; Prendergast & Belliveau, 2018), and health (Gray, 2011; Hassall & Balfour, 2016; Hodgins & Boydell, 2013; Westwood & Gordon, 2016). Drawing on the research-based theatre literature, we conceptualize theatre-based evaluation as an evaluation process rooted in the art of theatre.

A distinction should be made between arts-informed versus arts-based research (Cole & Knowles, 2008), and, similarly, between arts-based and arts-informed evaluation practices (Searle & Shulha, 2016). Whereas arts-informed research may use art as a component of a more traditional qualitative research design, for example, as a tool for data collection, arts-based work uses the chosen artform(s) to influence the entire research project. Research-based theatre encourages practitioners to consider how the various components of theatre might contribute to the inquiry process from start to finish. Although not all research-based theatre projects are participatory, many utilize collective creation, an approach to theatre-making in which a group creates a play collaboratively (Boal, 1992; Kaufman, 2001; Norris, 2000; Okello, 2016). The script-creation process can serve to collect and analyze participant voices in a similar way to other narrative-based qualitative approaches (Saldaña, 2003). Transforming participants' words into a dramatic form allows the grouping and connection of similar themes and experiences. Additionally, theatre gives space for the inclusion of emotion and physical gesture in the data-collection and -analysis process. Dramatic techniques are particularly well suited for capturing and conveying complex data, with the

ability to present multiple, even contradictory, narratives in coherent and engaging ways (Belliveau, 2006; Saldaña, 2003; Simons & McCormack, 2007). Whether relying on simpler techniques, such as selecting and juxtaposing excerpts of verbatim quotations that are read aloud with minimal theatrical accoutrement, or through the application of more advanced approaches, such as creative stage direction, fictionalization, and metaphor (Rossiter et al., 2008), theatre has a powerful potential to communicate diverse experiences.

Theatre as a methodology may be of particular interest to evaluators working from a utilization-focused perspective, with an emphasis on maximizing the likelihood that findings will translate into actions by the intended users of the evaluation (Patton, 2008). Creating, performing, and witnessing live theatre can be an emotional experience that may evoke strong responses in both actors and audiences (Cozart et al., 2003; Rossiter et al., 2008; Westwood & Gordon, 2016), conveying the emotional gravitas of participant experiences directly to stakeholders. Combined with theatre's immediacy and ephemeral nature, where a performance occurs as an interaction between actors and audience members in a particular moment in time, witnessing a play can powerfully motivate discussion, reflection, and action.

Other arguments for the use of theatre in evaluation include representing participant experiences more authentically and completely (Barone, 1991; Simons & McCormack, 2007); making participating in, and learning from, an evaluation more accessible to a wider audience (Rossiter et al., 2008; Simons & McCormack, 2007); and promoting ownership over findings by those whose experiences are being represented (Barone, 1991; Eakin & Endicott, 2006). Such a collaborative, empowering approach is particularly appropriate in communities like the DTES, where participants may feel disenfranchised and experience marginalization.

BEHIND THE SCENES: THE MAKING OF *VOICES UP!*

Figure 1 illustrates the relationship between the theatre-based evaluation project and the larger, more traditional program evaluation. *Voices UP!* was created over 22 months, during which time participants met regularly to devise, rehearse, perform, and refine the script. Although the larger program evaluation ended in August 2016, the play continued as participants revised and performed new versions of the script until late 2017.

The collective creation process occurred within the predefined framework of the Learning Lab program evaluation: the staff and program evaluator decided that the play would focus on community members' experiences of taking part in Learning Lab activities. The Learning Lab was running several ongoing clubs and workshops at the time, and each was to have its own scene in the play. The staff also wanted the process to move through the general phases of theatre production: script creation, rehearsal, and performance.

Community participants in the play were drawn from individuals currently participating in Learning Lab activities and reflected the diversity of the DTES.

Figure 1. Evaluation timeline

The three staff-member participants were Suzie O'Shea (S.O.S.), who ran the Learning Lab program; Andrew Leyland (A.L.), the program evaluator; and Christopher Cook (C.C.), the Learning Lab assistant and a theatre artist. Learning Lab staff and community members all had previously established relationships. Like all Learning Lab offerings, *Voices UP!* was “drop-in”—care was taken to inform all potential participants that no registration or commitment was required and that they were free to drop in or out at any point during the theatre-based evaluation process. The process outlined here flows from scripting to rehearsal to performance, and movement through these phases was determined by the group. In other words, the group collectively decided when it was time to move from scripting to rehearsal, and if and when the script would be presented to an audience. The process was non-linear, as new lines were created mid-rehearsal, and scenes were rewritten throughout the process.

Devising: Conversations to Dialogue

Chatting over tea and snacks—that is how the data collection for this theatre-based evaluation began. Seven community members and three staff participated in the initial script creation. Participants from the Learning Lab's Mahjong Club, an introductory art workshop, and a community button-blanket making project were invited to take part in hour-long conversations about their experiences. Some participants had past performance experience, but for many *Voices UP!* was their first play. As part of these conversations, C.C. introduced participants to research-based theatre and collective creation.

To make these initial conversations welcoming for participants who may have been uncomfortable with recording, the meetings were not audio or video

recorded. C.C. facilitated, asking questions such as “What was it like for you to take part in these activities?” and “What stands out to you when you think back on attending the workshop?” Based on notes of the responses, C.C. wrote the first draft of a scene and shared it with participants for revisions. Participants gathered weekly for one-hour meetings for the next eight weeks, conversing, writing, and revising the scenes. The resulting three scenes were framed by a collectively written prologue and epilogue, which knitted them together into a complete play. Participants were free to edit or withdraw their words from the script at any time and could choose whether they wanted to use a character name or their own name. Although attendance was not mandatory, absences were rare, and none of the participants dropped out of the script-creation process.

Rehearsal: From Page to Stage

All participants were given the option to perform their own words. Acknowledging that public speaking could present a barrier to participation, individuals could also record their dialogue for use during the performance or have another participant speak their words. Of the seven original community participants, five took part as live performers, one through audio recording and one by taking on the role of a director. An additional community member joined the cast as a voice actor, providing recorded monologues. All three staff members participated in the live performances.

As the collective brought the script “on its feet,” staging the performers’ movements, actions, and scene transitions, the script was considered fluid and was continually edited to reflect participants’ perspectives. If a participant wanted to change a line from their own story, they were free to do so. If a participant wanted to add a new line, the group came to a consensus by deciding whether the line fitted within the scope of the script (if it was related to their experience of the Learning Lab) and ensuring that other participants had the same opportunity to add dialogue. For example, when one participant wrote a line introducing themselves and describing their facilitator role at the Learning Exchange, all participants were given the option to add a similar line. Directing and staging decisions were facilitated by C.C. but were also made collectively. Rather than simply running through the script with participants reciting their lines, rehearsals were an extension of the devising process and took the shape of ongoing conversations.

Lights Up: Evaluation Takes the Stage

To create the most accessible performance situation possible, we presented *Voices UP!* as a stage-reading, with actors reading from scripts. For participants who found reading to be a barrier, fellow cast-members assisted with cuing and line recital.

Voices UP! was performed for the first time in May 2016, at a community–university academic conference hosted by the Learning Exchange in Vancouver’s DTES, to an audience of community members, researchers, and educators. The stage-reading was introduced to the audience by way of a pre-show speech, explaining the process used to create *Voices UP!* and preparing the audience for what

they were about to see. The live performance was intercut with audio-recorded monologues, dialogues, and projections, representing other voices from the program evaluation. The 25-minute performance was followed by a facilitated question and answer period (talk-back) between audience and participants that lasted nearly 45 minutes.

Encore Performances and Further Development

Successive performances over the next 18 months offered participants the chance to develop the script further. The second performance, presented in the Learning Exchange drop-in space for community members, local organizations, and UBC staff and students, included an added scene about an additional Learning Lab club, performed by two new participants. The script continued to be revised and developed for a third performance, which saw the addition of sock puppets, driven by several participants who expressed an interest in creating puppets for the show. Prior to the third performance, participants also designed and constructed a set. The third and fourth performances were presented to audiences of professional evaluators, at the 2017 Canadian Evaluation Society (CES) Conference, and the CES BC and Yukon Chapter Annual General Meeting, respectively.

Evaluation through Dramatization

The central image of the play is captured in its title, *Voices UP!*, and refers to the stories and perspectives of everyone involved in the project. The performance used a wide range of dramatic techniques, including puppets, audience participation, stage directions, and sound design, to bring the Learning Lab experience to life in ways that would be difficult to achieve with a written report.

The *Voice UP! script* (2017) contains four distinct scene categories. The first category, *Dramatizing Numbers*, facilitated the inclusion of quantitative data gathered as part of the larger program evaluation and background material necessary for the audience's understanding. The following script selection, from the play's prologue, provides an example of this scene category and draws on data such as attendance and facilitator records.

Prologue

SUZIE and ANDREW enter and address the audience.

SUZIE: (to audience) Over 500 individual workshop sessions, club meetings, and learning opportunities—

ANDREW: (to audience) Reaching nearly 400 local participants—

SUZIE: Made possible by over 40 volunteer facilitators—

ANDREW: Over the past four years—

SUZIE: Five days a week—

BOTH: Around 800 days of programming.

ANDREW: All of this makes up the Learning Lab at the UBC Learning Exchange.

SUZIE: My name's Suzie, I coordinate the Learning Lab. We support people to take a step outside their comfort zone, so we're standing at their learning edge together.

ANDREW: And I'm Andrew, I'm a UBC student, working at the Learning Exchange part-time, to evaluate the Learning Lab. It was a three-year pilot program—pretty experimental. And now it's been 3 years, so the question is: how'd we do?

BOTH: 500 workshop sessions, 400 participants, 40 facilitators.

ANDREW: The Learning Lab.

SUZIE: (to ANDREW) The Learning Lab is also ... much more than these numbers. It's the voices and stories of all the people who participate. To really understand the Learning Lab, you have to spend a day in the life with us. Ready, Andrew?

ANDREW: Let's go! (Voices UP! Collective, 2017)

The prologue also exemplifies performing self—these roles were written and performed by the staff members S.O.S and A.L., who portrayed themselves in the play and shared autobiographical stories. The prologue serves to invite audience members to explore “a day in the life” with the Learning Lab alongside A.L., and to witness both his journey as a program evaluator and the stories of the evaluation participants.

The second scene category, *Dramatizing Words*, drew on qualitative data collected as part of the larger program evaluation. For example, a monologue was created from notes that A.L. made after an informal interview with a member of the Mahjong Club. This community member was a part of the program evaluation but not a participant in the theatre-based evaluation project. Using these notes as a starting point, combined with the experiences of *Voices UP!* participants from the Mahjong Club, a composite character with the fictional name of Judy was created. Composite characters are drawn from the experiences of multiple individuals, encompassing more than one participant's story. Such characters offer privacy to individuals taking part in the project, allow for the grouping of similar experiences into a single character, and promote dramatization of the data. Not all data sets are inherently dramatic, and collected stories may need to be restoried into a new form to work as a monologue or dialogue on stage. Research-based theatre scholars have pointed to the tension between creating art and engaging in research, and to the pull for artistry to overtake the research (Belliveau, 2006; Saldaña, 1998, 2003). For theatre-based evaluators the same tension exists. Drama is the means by which the evaluation is being brought to life, not a licence for fictionalization. Nonetheless, some fiction, such as a composite character, has a place in theatre-based evaluation.

The third category of scene, *Dramatizing Conversations*, includes dialogue that was quoted from or inspired through conversations with participants during script devising and rehearsal. Participants engaged in these conversations after learning about collective creation, with the intention of creating lines that could

be used in the play. For example, in the play's Epilogue, a character named Maria offers the feedback, "We want activities that help us make connections in the community, to use the skills we're learning to give back" (Voices UP! Collective, 2017).

The fourth category, *Dramatizing Puppets*, includes scenes created for the sock puppets. Both the third and fourth categories provided a means to go beyond the larger program evaluation, collecting and conveying a deeper understanding of participants' experiences, such as the emotional and relational connections captured by a fourth selection from the script—the puppet scene. Puppets appeared to offer participants freedom from censorship (Cook & Belliveau, 2018). When devising the puppet scene, participants pulled on their sock puppets. Functioning like a mask, it was the puppets, not the participants, who were discussing the ups and downs of taking part in the Learning Lab. For example, the sock puppet Love says,

Don't underestimate relationship. We can talk-talk-talk-talk and it all comes back to one word: Relationship. ... How can you make the workshops and clubs ... more attractive? If you come up to me and say, not just—"Hey, this is happening now." But, "Hey, this is happening now and I'm going—do you want to come with me, do you want to come too?" (Voices UP! Collective, 2017)

PLAYMAKING PROCESS: GUIDING PRINCIPLES

Beyond traditional evaluation approaches, theatre-based evaluation offers a means of engaging and retaining participants, developing relationships, building capacity as both evaluators and theatre-makers, and stimulating in-depth discussions through performance. Theatre-based evaluation can also transform perspectives on evaluation. Rather than only an intellectual exercise, *Voices UP!* demonstrated that evaluation can be a creative, playful act, capable of emotionally involving participants and audiences through art-making.

We identify three core principles that guided our process:

- (1) *Creating collectively*. Just as "capacity-building is inherent to participatory approaches" to evaluation (MacLellan-Wright, Patten, dela Cruz, & Flaherty, 2007, p. 102), skill and knowledge sharing is intrinsic to collective creation. Group collaboration created a sense of shared ownership among participants and meant that no one person was required to be an expert at everything and everyone had a chance to lead. For example, while C.C. could assist with staging during rehearsals, he had no experience in set design or construction. Other participants led this process, teaching him and others how to saw and build a transportable stage set. The capacity building inherent in *Voices UP!* allowed participants to take part in an evaluation process in a creative manner, while also gaining transferable skills.

- (2) *Staging ourselves*. Staff worked alongside the participants, taking part in writing and performing, both as themselves and other characters. Embedding staff in the process allowed bonding and relationship building, as everyone shared the nervousness and excitement of each performance (Cook & Belliveau, 2018). The process created a collective identity, symbolized by the *Voices UP!* T-shirts that all participants wore to performances. As Chouinard (2013, p. 245) writes about participatory evaluation, "learning takes place" through "the relationships that are created." Staff and community participants supported each other while performing self, embodying personal stories on stage. Research in counselling psychology has suggested that performing self is an empowering experience that can catalyze personal growth and learning (Westwood & Gordon, 2016). In *Voices UP!*, performing self also ensured that self-reflexivity was present throughout the evaluation project (van Draanen, 2017), as participants devised and revised means of conveying their evolving experience to audiences over 22 months.
- (3) *We are all artists*. One of the challenges of using theatre in research and evaluation contexts is overcoming participants' fears of being inadequate artists (Simons & McCormack, 2007). *Voices UP!* took an approach of recognizing everyone's potential to contribute, regardless of experience or skill level in theatre. Fostering an environment where people felt welcome and comfortable, sometimes through something as simple as serving tea and apple slices during script-writing meetings and rehearsals, helped to make the prospect of engaging in artistic creation less intimidating. Allowing participants to choose their own roles and level of involvement supported community members and staff who were exploring theatre for the first time, while reflecting the Learning Lab's goal of providing low-barrier, strengths-based programming.

CURTAIN CALL: FINDINGS AND LESSONS LEARNED

Key findings of the larger Learning Lab program evaluation included the following: (a) sustainable activities grew from the patrons' interests; (b) activities focused on introductory art exploration were among the most popular; (c) activities that supported connections with external communities and organizations received positive responses from patrons as opportunities for them to give back. As illustrated by the excerpt from the Prologue, *Voices UP!* was able to communicate the numbers and findings captured in the final program-evaluation document to diverse audiences who otherwise may not have read the final report (Leyland, 2016).

The play also contributed new evaluation findings, through scenes such as the Epilogue, described above, and audience reaction. For example, during the second performance of the play, a Learning Exchange patron in the audience jumped up when A.L. spoke the line, "What did we forget [in our program evaluation]?" and

shouted a request for more one-on-one music lessons. After the performance, the patron shared their story of taking part in previous one-on-one Learning Lab music lessons, and their experience of high facilitator turnover. These are two examples of how evaluation data was generated directly from the use of theatre-based evaluation (Leyland, 2016).

Several lessons learned emerged from the *Voices UP!* creation and performance process that may be useful for others engaging in theatre-based evaluation:

- (1) *Theatrical expertise.* It was helpful to have at least one collective member with a theatre background to offer techniques, help draw out dramatic themes, and assist with writing and staging the performance. Groups interested in exploring theatre-based evaluation are advised to consult or collaborate with a theatre artist. For evaluators interested in learning more about collective approaches to theatrical exploration, see Norris (2000) and Johnstone (1987).
- (2) *Performances and talk-backs: a two-way street.* Performances offered a “two-way” experience, with the presence of a live audience creating a new context for learning (Cozart et al., 2003). Including opportunities for interaction and discussion after performances through talk-backs took advantage of this learning context. Talk-backs allowed participants to step into the evaluator role, asking questions of the audience about their experiences of the play and conversing with individuals from diverse communities.
- (3) *Facilitating an ending.* What would it be like if community members and staff were so excited by an evaluation process that they wanted to take part in it over and over again? During the summer of 2017, Learning Lab staff contemplated the legacy of the play once its performance run was over, while considering ways to create an empowering ending for those involved. After audience feedback at the third performance raised the idea of sharing the *Voices UP!* story in comic form, participants transitioned from collective play making to collective comic making. The cast created a comic book following the same collaborative process used to develop the play. Just as the character of Maria requests in the epilogue of *Voices UP!*, the comic creation process allowed participants to “use the skills” they learned “to give back” (Voices UP! Collective, 2017) and ensured that participants’ commitment of time, skills, thoughts, and feelings was recognized and celebrated in a lasting, tangible form. The comic is freely available online (Camman et al., 2017) for evaluators and community members who wish to learn more about the theatre-based evaluation.

EPILOGUE: FUTURE DIRECTION

Voices UP! illustrates what theatre can offer evaluation practitioners beyond traditional evaluation approaches. Script creation allows for a creative, participatory

means of data collection and analysis, one that has the potential to generate complex new insights through a broad range of theatre techniques, such as allowing participants to respond to questions using puppets. Performing a script created through theatre-based evaluation allows evaluators to share findings with diverse audiences in a more experiential, engaging way than a written report. As theatre-based evaluation allows participants to take leadership roles in both the evaluation process and the findings presentation, this participatory approach also creates an empowering learning experience.

This article shares one possibility for creating a theatre-based evaluation project, but any theatre paired with evaluation should be consistent with the program and context being evaluated. *Voices UP!* captured the stories of participants interested in taking part in a theatre project and is not meant to represent all voices from the Learning Exchange, or the DTES. Future research could explore the experiences of audience members who witness theatre-based evaluation projects (Belliveau & Nichols, 2017) or examine changes fostered and action taken after theatre-based evaluation projects are conducted. Our theatre-creation process began with script writing, but movement or voice explorations are also possible starting points for theatrical creation. Future research could explore theatre-based evaluation projects that engage in diverse forms of the theatre-creation process.

Although *Voices UP!* was a drop-in activity, many community members participated in the production for over a year, and six took part for all 22 months. *Voices UP!* connected participants to the evaluation process, involving them beyond isolated activities (such as interview and focus groups)—they belonged to a play, and to a collective built on relationships. This process was equally engaging for audience members, as *Voices UP!* played to packed houses and people were often turned away at the door. For community-based organizations for which evaluation is key but may be difficult to do consistently, effectively, and meaningfully, the *Voices UP!* process offers an example of an innovative way of bringing evaluation narratives and findings to life.

ACKNOWLEDGEMENTS

The excerpts from *Voices UP!* were written and performed by Jay Jay, Ken Lee, Andrew Leyland, Suzie O'Shea, Super Sky, and Teresa Shu-Tak Wong, with Chris Cook. The authors wish to thank all the participants, students, and staff who took part in *Voices UP!* collective creation process over its two-year duration, as well as the UBC Learning Exchange.

This research-based theatre project took place on the traditional territory of the Musqueam, Squamish, and Tsleil-Waututh Nations.

REFERENCES

- Balyasnikova, N., Higgins, S., & Hume, M. (2017). Enhancing teaching English as an additional language through playfulness: Seniors (Ethno)Drama Club in Vancouver's Downtown Eastside. *TESOL Journal*, 9(3), 481–497. <https://doi.org/10.1002/tesj.337>

- Barone, T. (1991). Assessment as theater: Staging an exposition. *Educational Leadership*, 48(5), 57. Retrieved from http://www.ascd.com/ASCD/pdf/journals/ed_lead/el_199102_barone.pdf
- Beck, J. L., Belliveau, G., Lea, G. W., & Wager, A. (2011). Delineating a spectrum of research-based theatre. *Qualitative Inquiry*, 17(8), 687–700. <https://doi.org/10.1177/1077800411415498>
- Belliveau, G. (2006). Engaging in drama: Using arts-based research to explore a social justice project in teacher education. *International Journal of Education & the Arts*, 7(5), 1–15. Retrieved from <http://www.ijea.org/v7n5/>
- Belliveau, G. (2007). Dramatizing the data: An ethnodramatic exploration of a playbuilding process. *Arts & Learning Research Journal*, 23(1), 31–51.
- Belliveau, G., & Lea, G. W. (Eds.). (2016). *Research-based theatre: An artistic methodology*. Bristol, England: Intellect.
- Belliveau, G., & Nichols, J. (2017). Audience responses to *Contact!Unload*: A Canadian research-based play about returning military veterans. *Cogent Arts & Humanities*, 4(1). <https://doi.org/10.1080/23311983.2017.1351704>
- Boal, A. (1992). *Games for actors and non-actors*. London, England: Routledge.
- Camman, C., Cook, C., Jay, J., Lee, K., Leyland, A., O'Shea, S., . . . & Shu-Tak Wong, T. (2017, November 23). *Voices UP! The comic*. Retrieved from <http://learningexchange.ubc.ca/2017/11/23/voices-up-the-comic/>
- Chouinard, J. A. (2013). The case for participatory evaluation in an era of accountability. *American Journal of Evaluation*, 34(2), 237–253. <https://doi.org/10.1177/1098214013478142>
- City of Vancouver. (2014). *Downtown eastside plan* (Local Area Plan). Vancouver, BC: City of Vancouver. Retrieved from <http://vancouver.ca/files/cov/downtown-eastside-plan.pdf>
- Cole, A., & Knowles, J. (2008). Arts-informed research. In J. G. Knowles & A. L. Cole (Eds.), *Handbook of the arts in qualitative research: Perspectives, methodologies, examples, and issues* (pp. 55–71). Thousand Oaks, CA: SAGE Publications. <https://doi.org/10.4135/9781452226545>
- Cook, C., & Belliveau, G. (2018). Community stories and growth through research-based theatre. *LEARNing Landscapes*, 11(2), 109–126. Retrieved from <https://www.learninglandscapes.ca/index.php/learnland/article/view/950>
- Cousins, J. B., & Earl, L. M. (1992). The case for participatory evaluation. *Educational Evaluation and Policy Analysis*, 14(4), 397–418. <https://doi.org/10.2307/1164283>
- Cousins, J. B., & Whitmore, E. (2007). Framing participatory evaluation. *New Directions for Evaluation*, 1998(80), 5–23. <https://doi.org/10.1002/ev.1114>
- Cozart, S. C., Gordon, J., Gunzenhauser, M. G., McKinney, M. B., & Petterson, J. A. (2003). Disrupting dialogue: Envisioning performance ethnography for research and evaluation. *Journal of Educational Foundations*, 17(2), 53.
- Donmoyer, R., & Yennie-Donmoyer, J. (1995). Data as drama: Reflections on the use of readers theater as a mode of qualitative data display. *Qualitative Inquiry*, 1(4), 402–428. <https://doi.org/10.1177/107780049500100403>
- Eagle Bear, D. (2016). *Gentrification and change* [Video series]. Vancouver, BC: UBC Learning Exchange, University of British Columbia. Retrieved from <http://learningexchange.ubc.ca/campus/students/student-orientation-materials/>

- Eakin, J. M., & Endicott, M. (2006). Knowledge translation through research-based theatre. *Healthcare Policy*, 2(2), 54–59. <https://doi.org/10.12927/hcpol.2007.18528>.
- Gillard, S., & Balyasnikova, N. (2015). *Fear, courage, and joy: Three conversations at the Learning Exchange*. Unpublished manuscript. Vancouver, BC: UBC Learning Exchange, University of British Columbia.
- Gray, J. (2011). *After the crash: A play about brain injury*. *Canadian Theatre Review*, 146(1), 66–86. <https://doi.org/10.1353/ctr.2011.0037>
- Hassall, L., & Balfour, M. (2016). Transitioning home: Research-based theatre with returning servicemen and their families. In G. Belliveau & G. W. Lea (Eds.), *Research-based theatre: An artistic methodology* (pp. 147–162). Bristol, England: Intellect.
- Hodgins, M. J., & Boydell, K. M. (2013). Interrogating ourselves: Reflections on arts-based health research. *Forum: Qualitative Social Research*, 15(1). <http://doi.org/10.17169/fqs-15.1.2018>.
- Johnstone, K. (1987). *Impro*. New York, NY: Routledge.
- Kaufman, M. (2001). *The Laramie project*. New York, NY: Vintage Books.
- Leyland, A. (2016). *Learning Lab evaluation report* [Unpublished report]. Vancouver, BC: UBC Learning Exchange, University of British Columbia.
- MacLellan-Wright, M. F., Patten, S., dela Cruz, A. M., & Flaherty, A. (2007). A participatory approach to the development of an evaluation framework: Process, pitfalls, and payoffs. *Canadian Journal of Program Evaluation*, 22(1), 99–124.
- Norris, J. (2000). Drama as research: Realizing the potential of drama in education as a research methodology. *Youth Theatre Journal*, 14(1), 40–51. <https://doi.org/10.1080/08929092.2000.10012516>
- Okello, B. F. (2016). The Ink Murderers can't hold it any longer. In G. Belliveau & G. W. Lea (Eds.), *Research-based theatre: An artistic methodology* (pp. 27–42). Bristol, England: Intellect.
- Patton, M. Q. (1994). Developmental evaluation. *Evaluation Practice*, 15(3), 311–319. [https://doi.org/10.1016/0886-1633\(94\)90026-4](https://doi.org/10.1016/0886-1633(94)90026-4)
- Patton, M. Q. (2008). *Utilization-focused evaluation*. Thousand Oaks, CA: Sage Publications.
- Prendergast, M., & Belliveau, G. (2018). Misperformance ethnography. In M. Cahnmann-Taylor & R. Siegesmund (Eds.), *Arts-based research in education: Foundations for practice* (pp. 99–114). New York, NY: Routledge.
- Rossiter, K., Kontos, P., Colantonio, A., Gilbert, J., Gray, J., & Keightley, M. (2008). Staging data: Theatre as a tool for analysis and knowledge transfer in health research. *Social Science & Medicine*, 66(1), 130–146. <https://doi.org/10.1016/j.socscimed.2007.07.021>.
- Saldaña, J. (1998). Ethical issues in an ethnographic performance text: The “dramatic impact” of “juicy stuff.” *Research in Drama Education: The Journal of Applied Theatre and Performance*, 3(2), 181–196. <https://doi.org/10.1080/1356978980030205>
- Saldaña, J. (2003). Dramatizing data: A primer. *Qualitative Inquiry*, 9(2), 218–236. <https://doi.org/10.1177/1077800402250932>
- Searle, M., & Shulha, L. (2016). Capturing the imagination: Arts-informed inquiry as a method in program evaluation. *Canadian Journal of Program Evaluation*, 31(1), 34–60. <https://doi.org/10.3138/cjpe.258>

- Simons, H., & McCormack, B. (2007). Integrating arts-based inquiry in evaluation methodology: Opportunities and challenges. *Qualitative Inquiry*, 13(2), 292–311. <https://doi.org/10.1177/1077800406295622>
- Towle, A., & Leahy, K. (2016). The Learning Exchange: A shared space for the University of British Columbia and Vancouver's Downtown Eastside communities. *Metropolitan Universities*, 27(3), 67–83. Retrieved from <https://eric.ed.gov/?id=EJ1120575>
- van Draanen, J. (2017). Introducing reflexivity to evaluation practice: An in-depth case study. *American Journal of Evaluation*, 38(3), 360–375. <https://doi.org/10.1177/1098214016668401>
- Voices UP! Collective. (2017). *Voices UP!* [Unpublished manuscript]. Vancouver, BC: UBC Learning Exchange, University of British Columbia.
- Westwood, M. J., & Gordon, A. (2016). Solace in theatre after war: Facilitating therapeutic change through theatre performance. In G. Belliveau & M. J. Westwood (Eds.), *Soldiers performing self in CONTACT!UNLOAD: Innovations in theatre and counselling psychology* (pp. 79–84). Vancouver, BC: UBC Faculty of Education.

AUTHOR INFORMATION

Christopher Cook worked at the UBC Learning Exchange as the Learning Lab assistant from 2015–2017. He recently completed his MA in counselling psychology at UBC and has worked as a theatre artist for the last ten years. He is passionate about exploring theatre as both a therapeutic, learning, and research tool. His previous playwriting credits include *The Better Parts of Mourning* (the frank theatre company), *Strip* (Staircase Theatre), *Gerty: Live! In Concert!* (BC Buds Festival), and *Quick Bright Things* (Persephone Theatre). *Voices UP!* offered Chris his first experience of collective creation.

Carolyn Camman is an evaluation specialist at the UBC Learning Exchange as well as an evaluation consultant working out of Vancouver, BC. Their involvement in *Voices UP!* came later in its process, first as an audience member who suggested turning the play into a comic book and later as an evaluator providing consultation support to that process, including conducting a literature review and helping prepare this article for publication.

Andrew Leyland is a graduate student in the School of Population and Public Health at UBC. His research involves intergenerational inequality and the policies that shape population health. He performed a program evaluation of the Learning Lab between 2015 and 2016. He is deeply grateful for the opportunity to undertake a community-based program evaluation at the UBC Learning Exchange, and to learn about theatre-based evaluation in the process. In his spare time, Andrew enjoys rock climbing and brewing beer.

Suzie O'Shea was the community animator at the UBC Learning Exchange from 2013–2018. She has a BA in Social Studies in Social Care, an MA in Youth and Community Studies, and was proud to be part of *Voices UP!* in her role as the coordinator of the Learning Lab program, exploring how arts-based approaches can build leadership and connections in an accessible way and significantly influence social change. *Voices UP!* was her acting debut, but Suzie has been found performing in contemporary line dancing, learning the ukulele, and trying her hand at sewing.

Angela Towle is the academic director of the UBC Learning Exchange. In this role she serves as the academic champion for the Learning Exchange, ensuring that it is linked

effectively with UBC's academic mission. Angela is an associate professor in the Department of Medicine and senior scholar in the Centre for Health Education Scholarship in the Faculty of Medicine at UBC. She is also Co-Director of Patient & Community Partnership for Education in the Office of UBC Health. Angela joined the *Voices UP!* team to facilitate the talk-back session after each performance.

APPENDIX: GLOSSARY OF THEATRE TERMS

arts-based research: Any research project that roots its methodological approach in creative expression or specific art forms.

audience participation: Inviting the audience to actively take part in the performance, such as by asking for volunteers to join a scene.

blocking: Performers' movements on stage, including entrance and exits cues and any physical gestures.

collective creation: A theatre project created by a group of people working together. The creators often take on dual roles as writer-performers, presenting the work themselves.

in-vivo dialogue: Script dialogue quoted directly from participants' own words.

performing self: Actors portraying themselves and telling their biographical stories onstage.

research-based theatre: A methodology in which theatre informs every aspect of the research project.

stage directions: In a script, the playwright's notes to the actors, directors, and other theatre artists regarding the set, lighting, sound, prop, costume, or projection design; character emotions, intentions, or blocking; intended audience experience; etc.

stage-reading: A performance in which actors are blocked, moving about the stage with their scripts in hand. Stage-readings may include design elements, such as sets and props. An alternative format is a script reading, in which actors read the script while seated or standing, and no design elements are present.

talk-back: Following the performance, a facilitated conversation between audience members and theatre-makers.

theatre-based evaluation: An evaluation project that draws on research-based theatre methodology to conduct an evaluation rooted in theatre creation.

Applying the Collaborative Approaches to Evaluation (CAE) Principles in an Educational Evaluation: Reflections from the Field

Jeremy Acree

University of North Carolina at Greensboro

Abstract: *This practice note presents reflections on the application of the collaborative approaches to evaluation (CAE) principles used as a guide in planning, implementing, and managing a collaborative evaluation in a higher educational setting. The reflection is based on the evaluation of a technology integration program intended to enhance K–12 teacher preparation in a school of education at a public university in the southeast United States. The evaluation was conducted during a one-year period by the author and a diverse team of novice and experienced evaluators. Discussion of the principles and their influence on collaborative practice are based on an analysis of evaluator reflections, meetings with stakeholders, and a culminating interview with stakeholders that were recorded and documented throughout the evaluation. Key takeaways from our reflection and analysis highlight the ways in which the CAE principles encourage reflection, the emphasis of some principles based on the specificities of context, and challenges applying the principles that emerged throughout the evaluation.*

Keywords: CAE principles, collaborative evaluation, education, evaluation practice, reflective practice

Résumé : *La présente note sur la pratique présente des réflexions sur l'application des principes visant à guider les approches collaboratives en évaluation (ACE) pour orienter la planification, la mise en œuvre et la gestion d'une évaluation collaborative dans un contexte d'enseignement supérieur. La réflexion est fondée sur l'évaluation d'un programme d'intégration de la technologie conçu pour la préparation des enseignantes et enseignants de la maternelle à la 12^e année, dans une faculté d'éducation d'une université publique du sud-est des États-Unis. L'évaluation a été réalisée pendant une période d'un an par l'auteur et par une équipe diversifiée d'évaluatrices et d'évaluateurs chevronnés et débutants. La discussion des principes et de leur influence sur la pratique collaborative découle d'une analyse des réflexions des évaluatrices et évaluateurs, de rencontres avec des intervenants et d'une entrevue finale avec les intervenants, qui ont été enregistrées et consignées tout au long de l'évaluation. Les principales conclusions tirées de notre réflexion et de notre analyse*

Corresponding author: Jeremy Acree, University of North Carolina at Greensboro, 1300 Spring Garden St, Greensboro, NC 27412, USA; jgacree@uncg.edu

portent sur les façons dont les ACE encouragent la réflexion; l'accent mis sur certains principes selon le contexte; et les défis liés à l'application des principes qui ont été notés pendant l'évaluation.

Mots clé : principes d'ACE, évaluation collaborative, éducation, pratiques d'évaluation, pratique réfléchie

Collaborating with stakeholders in the planning and implementation of evaluation involves navigating and managing expectations, communication, and relationships (King, 2005). These uncertain dimensions of practice require that evaluators not only draw from theoretical knowledge but also conduct what Schon (1983) refers to as “reflection-in-action” in order to make thoughtful decisions based on experience, inclinations, and predispositions (Chouinard et al., 2017). Designing and conducting collaborative evaluation thus requires flexibility in methods, models, and techniques (Cousins, Whitmore, & Shulha, 2013). Acknowledging the inherent contingencies in this type of approach, Shulha, Whitmore, Cousins, Gilbert, and al Hudib (2016) developed the collaborative approaches to evaluation (CAE) principles to offer empirically derived practical guidance while allowing space for evaluators to respond to the specificities of context. As Shulha et al. explain, the principles offer a “system for thinking about collaborative approaches to evaluation (CAE) ... intended to support considerations of professional practice, both generally and in situ” (p. 193).

The CAE principles are a product of extensive empirical analysis and synthesis. Through a sequence of both in-depth and broad studies, the perspectives and experiences of more than 300 practicing evaluators were distilled into eight interconnected principles, revealing a comprehensive depiction of collaborative approaches in practice (Shulha et al., 2016). Despite the substantial time and effort spent in their development, Shulha et al. (2016) assert that the principles should be viewed as fluid and subject to ongoing revision. They suggest that the utility of the principles will be clarified and refined through further empirical study as evaluators reflect upon and analyze their application in practice.

In this practice note, I reflect on the experience of working with the CAE principles as a guide in planning, implementing, and managing a year-long collaborative inquiry in a higher education setting. This analysis connects directly to two of the four suggestions provided by Shulha et al. (2016) for integrating the principles into evaluation practice and scholarship: (1) applying the principles as “a guide to planning and implementing CAE,” and (2) using the principles as “a basis for retrospective reflection on completed projects (with an eye to surfacing lessons learned)” (p. 212). I used the following questions to guide my analysis of the CAE principles:

- In what ways and to what extent do the principles influence and guide practice throughout the evaluation process?
- In what ways are the principles useful to novice evaluators?

To address these questions, I analyzed evaluator reflections, meetings with stakeholders, and a culminating interview with stakeholders. The practice note begins with an overview of collaborative approaches to evaluation and the CAE principles, followed by a description of the context of the evaluation. It concludes with a discussion of lessons learned regarding collaborative approaches to evaluation and the utility of the principles to guide practice.

COLLABORATIVE APPROACHES TO EVALUATION

The CAE principles are not defined by a singular technique or method but refer to a family of approaches which call for a partnership to conduct systematic inquiry between evaluators who understand evaluation processes and methods and those communities, individuals, or groups with knowledge of the program and community context (Cousins & Chouinard, 2012). Numerous authors (see Cornwall & Jewkes, 1995; Cousins & Chouinard, 2012; Cousins et al., 2013) have provided lists of approaches that fit under the collaborative umbrella, including participatory (e.g., Cousins & Whitmore, 1998; King, 2005), transformative (e.g., Cousins & Whitmore, 1998; Mertens, 2009), empowerment (Fetterman, 2001), collaborative (O'Sullivan, 2004, 2012; Rodríguez-Campos, 2005, 2012), and developmental evaluation (Patton, 2011), among others. Throughout the literature, the term “collaborative” is used loosely and often interchangeably with “participatory,” reflecting the broad nature of this categorization.

The CAE principles are heavily influenced by participatory research methods, such as participatory action research and rapid rural appraisal, which Cousins et al. (2013, p. 8) suggest were “developed at least in part as a reaction to positivist models of inquiry that were seen as exploitive and detached from urgent social and economic problems.” Collaborative approaches emphasize the importance of context and often rely on stakeholder involvement to generate findings that are relevant and useful for local problem solving (Cousins & Whitmore, 1998). According to Cornwall and Jewkes (1995, p.1668), participatory methodologies draw attention to “who defines research problems and who generates, analyses, represents, owns and acts on the information which is sought.” Participatory research goes beyond the inclusion of program participants as data sources, addressing issues of control over the research agenda and process.

Cousins and Chouinard (2012) specify three justifications for collaborative inquiry: pragmatic, political, and philosophical. Approaches based on a pragmatic justification are oriented toward problem solving, aiding organizational decision making, and facilitating the use of evaluation findings and processes (Cousins & Whitmore, 1998). Politically motivated approaches focus on social action and transformation through the empowerment of individuals and groups, particularly those who have been historically marginalized (Cousins & Whitmore, 1998). Although these two streams of participatory evaluation differ in their primary emphasis, they share a philosophical justification in that they aim to produce deeper, more meaningful knowledge through collaboration (Cousins et al., 2013).

Empirical studies have found that collaborative approaches to evaluation can increase the utilization of evaluation findings and processes and can promote individual and organizational learning (Cousins & Chouinard, 2012). By taking an educative role, evaluators can foster intra-organizational dialogue and a culture of evaluative thinking (Cousins & Earl, 1992). However, these instrumental and conceptual benefits are not without limitations. Cornwall and Jewkes (1995) highlight the complexity of participatory methods and the challenges that can arise from unintended consequences, conflicting agendas within communities, and practical limitations of time and resources. Stakeholder needs, predispositions, and levels of commitment to evaluation influence the degree to which evaluation can achieve political and pragmatic goals (Cousins et al., 2013).

THE CAE PRINCIPLES

The CAE principles serve as a pragmatic tool for navigating the partnership between evaluators and program and community stakeholders across unique and complex contexts (Shulha et al., 2016). The authors of the principles do not specify how they are to be applied in practice, but they do provide two considerations regarding their utility. First, they suggest that each principle is essential, and implementation of each is a matter of degree, not a “whether-or-not” decision. Second, they assert that the principles carry only a “loose temporal order” and that certain principles should not be emphasized a priori. The eight principles are presented as follows:

- clarify motivation for collaboration,
- follow through to realize use,
- promote evaluative thinking,
- monitor evaluation progress and quality,
- monitor and respond to the resource availability,
- promote appropriate participatory processes,
- develop a shared understanding of the program, and
- foster meaningful relationships. (Shulha et al., 2016, p. 194)

Despite the recency of their publication, some research already exists on the use of the principles in practice. A forthcoming book, *Global Test Drive of Principles for Collaborative Approaches to Evaluation (CAE)*, edited by J.B. Cousins, presents empirical studies of the application of the principles in practical and pedagogical settings (Cousins, forthcoming). In one chapter, Searle, Merchant, Chalas, and Lam (previously published in 2017) describe using the principles as a framework for examining how evaluation theory connects with practice. In a case study of a developmental evaluation, they found that the principles “allow the reflective evaluator to focus on the elements of collaboration, allowing her to better understand the challenges and success of the evaluation” (p. 368). In another chapter, Chouinard, Cavanaugh, Adetogun, and Baker (forthcoming) describe

their application of the principles as a pedagogical tool, engaging students in dialogue and facilitating reflection on the purpose, methods, and consequences of collaborative practice. In the subsequent analysis, I present further evidence regarding the utility of the principles as a guide to practice and tool for reflection and learning.

CONTEXT

The Transforming Teaching through Technology (T4) program is a wide-ranging initiative supporting the School of Education at a public university in the south-east United States and its partnership with seven K–12 schools in two nearby districts. The project is focused on integrating emerging technologies into the institution's teacher preparation program curriculum and local classroom instruction by providing monetary incentives, technology resources, and support staff. Components of T4 include hands-on learning studios known as “makerspaces” at the university and seven partner schools, summer professional development camps for K–12 teachers, and working groups of university faculty focused on technology integration in teacher preparation courses. At the time of the evaluation examined here, the program was in its fourth year of a five-year grant cycle. Evaluations in the first three years focused on development and improvement of individual program components.

The evaluation team consisted of three graduate students (including the author), a faculty advisor, and an external evaluator. We applied a mixed method design to capture both impact and sustainability data from multiple sources, including faculty, pre-service students, grant support staff, and program documentation. The primary mode of collaboration was formal meetings with a stakeholder group assembled by the evaluation team with assistance from T4's principal investigator. The group comprised partner-school liaisons, university-based program staff, and faculty with close ties to the program.

While we would have preferred a more collaborative approach with our stakeholder group, due to their busy schedules we were able to collaborate with them only during specific points in the evaluation. We held three formal meetings, one each during the design phase of the evaluation, prior to data collection, and after an initial round of data analysis. Each meeting was roughly three hours in length and structured around sharing updates on the evaluation with stakeholders, facilitating dialogue on future directions, and providing a forum for stakeholder input, questions, suggestions, and clarifications. When stakeholders could not attend formal meetings, we followed up with individual conversations over the phone, through email, or in person.

PRACTICAL SIGNIFICANCE AND LESSONS LEARNED

To explore and test the utility of the CAE principles, we selected specific points during the planning, implementation, and reporting of the evaluation to review

the principles, question our understanding of them, and reflect on the extent to which certain principles influenced our practice. In our discussions, the evaluation team engaged in what [Buckley, Archibald, Hargraves, and Trochim \(2015\)](#) call “evaluative thinking,” critically reflecting on the meaning and relevance of the principles in our collaborative work. Through this dialogue, we surfaced differing perspectives and experiences, examined individual and collective assumptions, and converged on points of consensus ([Chouinard et al., forthcoming](#)).

In this practice note, my analysis of the influence and utility of the CAE principles was based on documentation of evaluation team meetings, reflective discussions, and stakeholder group meetings through note taking and audio recordings, which were then selectively transcribed. A brief group interview with stakeholders at the end of the evaluation was also audio recorded and transcribed. Those documents served as the primary data sources for the key takeaways presented below, although informal conversations with evaluation team members have also shaped some of the findings and discussion. The opportunities, benefits, and challenges that emerged from our analysis are discussed below.

Principled Reflection

Using the CAE principles as a guide led us to critically reflect throughout the evaluation on the methods and complexities of collaborative inquiry. A consistent tension between the flexible nature of the principles and the situational demands of practice stimulated dialogue among the evaluation team regarding which principles to emphasize, when, and in what ways. [Shulha et al. \(2016\)](#) intentionally avoid prescriptive direction, and that ambiguity turned our application of the principles into an ongoing, negotiated process among the evaluation team. How can we best foster meaningful relationships with stakeholders? How do we promote evaluative thinking and what does it look like when it occurs? Is our evaluation likely to be used? Questions such as these were raised around many of the principles as we deliberated on the evaluation approach and methods.

Reflection was particularly beneficial for the team’s three novice evaluators, who had no prior experience conducting collaborative evaluation. As noted by one evaluator in a team meeting, “this discussion [of principles] helps us to identify areas where we need further clarification regarding our evaluation plan and about the program.” Taking the time to discuss how the principles applied to specific methods and decisions clarified expectations for collaboration and led to thoughtful and intentional discussions with stakeholders. For example, reflections after our first meeting with stakeholders generated a list of discussion topics for our second meeting, focused on revising initial plans as well as outlining specific next steps. The nature of the principles encourages—and perhaps requires—reflection if they are to be applied meaningfully in practice. For us, that reflection ultimately led to deeper understanding of collaborative inquiry and practical insights into the program and our approach to the evaluation.

Prioritization of Principles

Over the course of the evaluation, some principles received more consideration than others. Although we sought to map our practice at different points onto each of the principles, the broad scope and compartmentalized nature of T4 led us to focus our dialogue extensively on three principles: clarifying motivation for collaboration, developing a shared understanding of the program, and following through to realize use.

We initially expected that *clarifying motivation for collaboration* would be a product of conversations with stakeholders during the early stages of the evaluation, laying the foundation for the remainder of our work together. Instead, this principle was relevant throughout the evaluation and was a topic of discussion from the design phase until after our final report was sent to the project director. Articulating the purpose of collaborative evaluation and establishing expectations with stakeholders was an ongoing process. A presentation on CAE in our first meeting with stakeholders initiated dialogue, but interviews at the end of the evaluation indicated that some uncertainty was never fully resolved. One stakeholder felt he contributed to the evaluation, saying, “I was on the ground, dealing with the teachers and students in the schools, so I thought I was able to provide some insight in that arena.” Others, though, felt that meaningful collaboration was limited. One stakeholder said, “evaluators have come in over the years, watching us, but I don’t feel like I have given much input into the evaluation.” Shulha et al. (2016, p. 200) suggest that “not only is it important to establish the meaning of the CAE application early, there are also benefits to reinforcing this meaning over time.” This was true for us, and additional discussion around the purpose of collaboration probably would have improved multiple aspects of the evaluation.

The organizational structure of T4 made *developing a shared understanding of the program* an important principle for our collaboration with stakeholders. University faculty, program staff, partner-school liaisons, and even some graduate students each have critical, but often isolated, roles in T4. Our stakeholder meetings brought representatives from these groups together, and discussions of how the various components of T4 fit together and worked toward a common aim was beneficial for both stakeholders and evaluators. As one stakeholder said early in the year, “I’d like to get an idea of what’s happening in other places. ... I just get a small view of it. I’m sure there’s a lot more going on.” Stakeholder perspectives also informed our decision to focus on program sustainability, and their input was a primary factor in defining the indicators we used to guide data collection and analysis. Shulha et al. (2016) emphasize responsiveness to context and stakeholder needs throughout their description of the principles, and a focus on local, situational demands is particularly evident in *developing a shared understanding of the program*. In the T4 evaluation, that emphasis pushed us to challenge our own comprehension of the program goals and activities and to integrate the perspectives of stakeholders to form a more comprehensive understanding.

In the first three years of the T4 program, it was difficult to identify examples of how our evaluations were being used in program decisions or used as a source

of learning and program improvement. Those evaluations were intended primarily for the principal investigator and were included in federal reporting documents but were rarely a topic of program-level discussions. Therefore, one motivation for adopting a collaborative approach in the fourth year was a desire to increase both instrumental and conceptual uses of evaluation. By bringing together a diverse group of stakeholders, we could share the evaluation process and findings directly with individuals who could then incorporate the knowledge we built collectively into their own work. The CAE principles encouraged us to focus on practical outcomes of evaluation at the individual level, which Shulha et al. (2016, p. 210) describe as changes “in disposition toward the program or evaluation, and the development of program skills including systematic evaluative inquiry.” By the end of our evaluation, stakeholders expressed enthusiasm for the knowledge we produced and a better understanding of the process and potential of evaluation. As one stakeholder said, “I am thinking now, is what I’m doing working? I guess that’s what you’re doing all the time [as evaluators], figuring out if what you’re doing is working. . . . I just want to see successes.” Stakeholders also expressed interest in continued participation in evaluation in future years.

Limitations and Challenges

Many of the challenges we faced in the evaluation of T4 and in our collaboration with stakeholders were a function of uncertainty around the commitment to and clarity of collaborative inquiry. As novice evaluators, we struggled to define expectations and structure collaboration in a way that would be optimal for stakeholders and evaluators. As a result, some stakeholders were active participants in meetings throughout the year, while attendance for others was inconsistent. The history and context of the evaluation may also have limited the depth of participation. Familiarity between stakeholders and evaluators built over the first three years of the program was beneficial in soliciting initial buy-in, but this was the first year we had adopted a collaborative approach, requiring more engagement from stakeholders than in previous years. Participation peaked and waned, and it was often difficult to facilitate meaningful collaboration. Literature on collaborative evaluation and the CAE principles made us aware of some of the challenges of collaborative inquiry, but it also cautioned that such inquiry is contingent and rarely follows a script (Cousins & Chouinard, 2012). As Shulha et al. (2016, p. 203) state, “whether or not evaluators have the wherewithal to mitigate the erosion of commitment to CAE will depend both on the skills of the evaluator and emerging conditions within the organization.” The principles were beneficial in surfacing issues and drawing our attention to key dimensions of collaborative inquiry, but success in this type of approach is not achieved by following a prescribed set of directions. What became quite clear to us is that collaborative evaluation requires a blend of theoretical knowledge and practical experience, strong relationships, a receptive organizational culture, and commitment from stakeholders and evaluators.

CONCLUSION

Shulha et al. (2016, p. 196) suggest that “the value of [the CAE] principles will rest in their capacity to illuminate complexity rather than resolve it, to inform decisions rather than prescribe them.” In utilizing the principles as a guide to collaborative practice, our evaluation team gained a better understanding of the nuances of the program we evaluated and the dynamic nature of soliciting, managing, and integrating stakeholder participation into evaluation practice. The principles also served as a teaching tool for novice evaluators. We gained a better understanding of the potential depth and scope of collaboration with stakeholders and learned through consistent reflection how to shape our evaluation approach to fit the program context. Use of the CAE principles as a guide to practice did not alleviate the ambiguity and uncertainty of collaboration but was beneficial in initiating discussions, clarifying motivations and aims, and encouraging practices that promoted evaluation use.

REFERENCES

- Buckley, J., Archibald, T., Hargraves, M., & Trochim, W. M. (2015). Defining and teaching evaluative thinking: Insights from research on critical thinking. *American Journal of Evaluation*, 36(3), 375–388. <https://doi.org/10.1177/1098214015581706>
- Chouinard, J. A., Boyce, A. S., Hicks, J., Jones, J., Long, J., Pitts, R., & Stockdale, M. (2017). Navigating theory and practice through evaluation fieldwork: Experiences of novice evaluation practitioners. *American Journal of Evaluation*, 38(4), 493–506. <https://doi.org/10.1177/1098214016667582>
- Chouinard, J. A., Cavanaugh, S. A., Adetogun, A., & Baker, K. (Forthcoming). A view from the classroom: A reflection on the use of the CAE principles in an instructional setting. In J. B. Cousins (Ed.), *Global test drive of principles for collaborative approaches to evaluation (CAE)*. Thousand Oaks, CA: SAGE.
- Cornwall, A., & Jewkes, R. (1995). What is participatory research? *Social Science Medicine*, 41(12), 1667–1676. [https://doi.org/10.1016/0277-9536\(95\)00127-s](https://doi.org/10.1016/0277-9536(95)00127-s)
- Cousins, J. B. (Ed.). (Forthcoming). *Global test drive of principles for collaborative approaches to evaluation (CAE)*. Thousand Oaks, CA: SAGE.
- Cousins, J. B., & Chouinard, J. A. (2012). *Participatory evaluation up close*. Charlotte, NC: Information Age Publishing, Inc.
- Cousins, J. B., & Earl, L. M. (1992). The case for participatory evaluation. *Educational Evaluation and Policy Analysis*, 14(2), 397–418. <https://doi.org/10.2307/1164283>
- Cousins, J. B., & Whitmore, E. (1998). Framing participatory evaluation. *New Directions for Evaluation*, 1998(80), 5–23. <https://doi.org/10.1002/ev.1114>
- Cousins, J. B., Whitmore, E., & Shulha, L. (2013). Arguments for a common set of principles for collaborative inquiry in evaluation. *American Journal of Evaluation*, 34(1), 7–22. <https://doi.org/10.1177/1098214012464037>

- Fetterman, D. M. (2001). *Foundations of empowerment evaluation*. Thousand Oaks, CA: SAGE.
- King, J. A. (2005). Participatory evaluation. In S. Mathison (Ed.), *Encyclopedia of evaluation* (pp. 292–296). Thousand Oaks, CA: Sage. <https://doi.org/10.4135/9781412950558>
- Mertens, D. M. (2009). *Transformative research and evaluation*. New York, NY: Guilford Press.
- O’Sullivan, R. (Ed.). (2004). *Practicing evaluation: A collaborative approach*. Thousand Oaks, CA: SAGE.
- O’Sullivan, R. (2012). Collaborative Evaluation within a framework of stakeholder-oriented evaluation approaches. *Evaluation and Program Planning*, 35(4), 518–522. <https://doi.org/10.1016/j.evalprogplan.2011.12.005>. Medline: 22364849
- Patton, M. Q. (2011). *Developmental evaluation: Applying complexity concepts to enhance innovation and use*. New York, NY: Guilford Press.
- Rodríguez-Campos, L. (2005). *Collaborative evaluations: A step-by-step model for the evaluator*. Tamarac, FL: Llumina Press.
- Rodríguez-Campos, L. (2012). Advances in collaborative evaluation. *Evaluation and Program Planning*, 35(4), 523–528. <https://doi.org/10.1016/j.evalprogplan.2011.12.006>. Medline: 22377419
- Schon, D. (1983). *The reflective practitioner: How professionals think in action*. New York, NY: Routledge.
- Searle, M. J., Merchant, S., Chalas, A., & Lam, C. Y. (2017). A case study of the guiding principles for collaborative approaches to evaluation in a developmental evaluation context. *Canadian Journal of Program Evaluation*, 31(3), 350–373. <https://doi.org/10.3138/cjpe.328>
- Shulha, L. M., Whitmore, E., Cousins, J. B., Gilbert, N., & al Hudib, H. (2016). Introducing evidence-based principles to guide collaborative approaches to evaluation: Results of an empirical process. *American Journal of Evaluation*, 37(2), 193–215. <https://doi.org/10.1177/1098214015615230>

AUTHOR INFORMATION

Jeremy Acree is a PhD student in the department of Educational Research Methodology at the University of North Carolina at Greensboro. He works as an evaluator in the university’s Office of Assessment, Evaluation, and Research Services.

Démocratisation et efficience sont-elles antinomiques au niveau local?

Le cas des réformes managériales des administrations publiques locales en Wallonie¹

Catherine Fallon
Université de Liège

Résumé : *La généralisation du recours aux instruments du « New Public Management » (NPM), au nom de l'efficacité et de l'efficience de l'action publique, impose de reformuler la question du lien entre démocratie et administration. Au niveau local, il est possible d'analyser à nouveau frais le triangle entre politique – administration – citoyens, tel qu'il évolue suite à l'appropriation des nouveaux instruments de gestion stratégique et des dispositifs participatifs. L'étude de l'émergence de ces dispositifs en Wallonie met en évidence de nouveaux modes d'exercice du pouvoir et d'action publique et pose la question de leur inscription dans une logique de renforcement démocratique.*

Mots clé : *Administration publique, gestion stratégique, participation, pouvoirs locaux*

Abstract: *The generalized use of new public management tools, in the name of the efficiency and effectiveness of public action, forces us to question once more the relationship between democracy and administration. At a local level, it is possible to reanalyze the politics-administration-citizens triangle, as it evolves following the appropriation of new strategic management tools and participation mechanisms. The study of the emergence of these new mechanisms in the Walloon region of Belgium highlights new means of exercising power and public action and bring into question their involvement in the logic of democratic reinforcement.*

Keywords: *public administration, strategic management, participation, local government*

La généralisation du recours aux instruments du « New Public Management » (NPM), au nom de l'efficacité, de l'efficience et/ou de la transparence de l'action publique, impose de reformuler la question du lien entre démocratie et

Correspondance à l'auteur : Catherine Fallon, SPIRAL, Département de Sciences politiques de l'Université de Liège; Catherine.fallon@ulg.ac.be

administration (Fallon et Ficet, 2017). Que devient aujourd'hui le lien entre citoyens et autorités politico-administratives? Cette question est pertinente dans un contexte où l'espace politique est de plus en plus complexe et technocratique et où le contrôle des citoyens sur les autorités par la voie représentative est aujourd'hui souvent contesté ou largement considéré comme défaillant. Le lien entre citoyens et autorité politico-administratives peut-il être renforcé par les nouveaux dispositifs de responsabilisation de l'administration ou encore par la participation du public?

Au niveau local, il est possible d'analyser à nouveau frais le triangle entre politique – administration – citoyens, tel qu'il évolue suite à l'appropriation des nouveaux outils de gestion publique et de mécanismes innovants de participation des citoyens et des usagers. Cet espace politico-administratif fait l'objet de peu d'études spécialisées en matière d'évaluation, alors que c'est à ce niveau qu'il est plus facile de développer des approches transversales de gestion publique sur un territoire limité ou de tirer parti d'une proximité plus grande entre autorités publiques, administration et citoyens. Cette contribution propose une analyse critique du déploiement de nouveaux dispositifs qui émergent dans de nombreux secteurs de l'action publique locale en Wallonie (Fallon, Hansoul et Piron, 2016). Après un court rappel des principes de gestion publique propres au courant européen du NPM, cet article analyse en détail la mise en œuvre d'un nouvel instrument de programmation stratégique que mobilisent les autorités publiques en Wallonie depuis quelques années, le PST ou Programme stratégique transversal, pour le comparer à d'autres dispositifs emblématiques d'une approche participative plutôt que gestionnaire, propres aux politiques de développement local. Le travail présente une comparaison entre les qualités de ces instruments mobilisés par les gestionnaires locaux et les principes de transformation de la gestion publique vers un modèle « neo-weberien » efficace et ouvert aux usagers (Pollitt et Bouckaert, 2011).

DÉMOCRATIE ET NEW PUBLIC MANAGEMENT

Dans leur ouvrage de référence comparant les réformes des services publics engagées dans douze États occidentaux ainsi qu'au sein de la Commission européenne, Pollitt et Bouckaert (2011) isolent plusieurs variables au cœur des trajectoires de réforme des modes de gestion publique inspirées par ce courant : la gestion financière, la gestion du personnel, les formes organisationnelles, la mesure de la performance et la transparence. Les réformes en matière de finances locales transforment le budget en outil de gestion et de planification aux mains des gestionnaires publics : son organisation en structure à double entrée, proche des instruments comptables propres au secteur privé, facilite la meilleure communication publique et une meilleure reddition des comptes pour les programmes politiques ciblés. Au niveau organisationnel, le courant de réforme cherche à résoudre la tension entre une volonté de spécialisation des fonctionnaires et un souci de coordination entre les services. Les discours prônent encore une

« responsabilisation » des agents dans la gestion des projets, une plus grande orientation vers les résultats plutôt que sur le respect des procédures. Ils encouragent aussi l'introduction de systèmes de mesure de la performance, censés assurer un pilotage efficace de l'action publique : ces mesures tendent à focaliser l'attention sur les extrants, qui sont les réalisations directes des agents, plutôt que sur les résultats auprès des bénéficiaires, plus difficiles à estimer. Ces réformes cherchent à favoriser une meilleure efficacité de l'action publique et une communication plus transparente entre administration, responsables politiques et citoyens. Elles transforment, comme une forme de pouvoir disciplinaire, la culture organisationnelle en imposant d'autres cadrages pour l'action publique et sa problématisation (Chouinard et Milley, 2015).

Ces principes gestionnaires doivent être mis en perspective avec les qualités propres à la démocratie libérale, mais les domaines renvoient à deux ordres normatifs différents : le premier à une rationalité instrumentale orientée vers l'efficacité et l'efficience et le second à une rationalité en valeur, encadrée par le processus électif de sélection des représentants du peuple. Une approche trop managériale tend à concentrer l'attention de l'administration sur l'efficacité de son action, quitte à négliger la dimension collective de sa mission de service public en marginalisant la parole des citoyens et des publics concernés. Certaines dimensions du NPM peuvent néanmoins contribuer à rapprocher l'action de l'État du citoyen, venant pallier les imperfections des régimes représentatifs par des mécanismes de compte rendu et d'évaluation ainsi que des dispositifs de participation directe. On retrouve les deux principes de transformation que soulignaient [Pollitt et Bouckaert \(2011\)](#) dans leur présentation de l'état « neo-weberien » : une tendance à l'intégration des principes du managérialisme, pour plus d'efficacité et d'efficience avec une augmentation de l'offre de participation des citoyens.

Une démarche évaluative sous le seul contrôle de l'administration contribue indirectement à renforcer la structure bureaucratique elle-même ([Dahler-Larsen, 2012](#)) tandis que les processus participatifs ou parlementaires, plus ouverts, renforcent son potentiel démocratique ([Weiss, 1999](#)). Dans son analyse de la tension entre crise démocratique et nouvelles modalités de gestion publique, [Ficet \(2017\)](#) propose pour sa part de se référer à trois traits emblématiques des démocraties libérales contemporaines : l'implication des citoyens à l'élaboration des normes, le contrôle de l'autorité politique et celui de l'arbitraire potentiel des administrations.

Cette contribution présente une analyse de cette tension entre efficacité gestionnaire et démocratique au niveau de l'action publique locale, niveau de pouvoir trop souvent délaissé par les analystes politiques. C'est à ce niveau qu'il est intéressant d'analyser les dynamiques de transformation des relations entre élus, fonctionnaires et citoyens. Dans un souci d'approche pragmatique, l'analyse prend appui sur le concept d'instrumentation et d'instrument d'action publique (IAP) compris comme un « dispositif socio-technique qui organise des rapports sociaux spécifiques, entre puissance publique et destinataires, qui sont porteurs de valeurs, nourris d'une interprétation du social, et d'une conception précise du mode de régulation envisagé » ([Lascoumes et Le Galès, 2004](#), p. 13). Les

auteurs soulignent l'importance d'analyser, au-delà des fonctions techniques de l'instrument, la définition politique que traduit celui-ci quant à la relation entre gouvernants et gouvernés. Trois dispositifs récents, qualifiés de managériaux, stratégiques ou participatifs, ont été analysés : ces instruments sont présentés dans leur matérialité comme des marqueurs des mutations de l'exercice légitime du pouvoir politique, parce que loin d'être neutres, ils traduisent dans leur conception et leur fonctionnement un ensemble de valeurs politiques ainsi qu'un rapport spécifique entre l'État et la société. En analysant au plus près du terrain, le développement de nouvelles pratiques, des discours et valeurs qui les sous-tendent et les légitiment, les stratégies et la rhétorique des acteurs, ainsi que la transformation des cadres organisationnels, il est possible de définir pour chaque instrument les tensions spécifiques entre les deux pôles, techniques et politiques. L'analyse de l'instrumentation permet de comparer les caractéristiques des instruments en fonction de leur période d'émergence et de proposer des hypothèses quant aux lignes directrices d'évolutions futures, au niveau local dans notre cas.

Les deux questions de recherche que nous proposons d'aborder deviennent alors : 1) voit-on émerger de nouveaux modes d'exercice du pouvoir politique et administratif du fait de la généralisation de dispositifs de pilotage et d'évaluation de l'action publique au niveau local et 2) dans quelle mesure cette nouvelle instrumentation s'inscrit-elle dans une logique participative et démocratique?

Le matériel empirique mobilisé pour cette étude résulte de travaux récents menés par l'auteur et son équipe en ce qui concerne les pouvoirs locaux wallons. Une étude de cas a été réalisée par l'auteur sur une longue période (2013–2016) dans une commune rurale de Wallonie², laquelle avait mis en place de façon volontaire un PCS (Plan de cohésion sociale), un PCDR (Programme communal de développement rural) et un PST (Programme stratégique transversal) : elle a permis de rassembler des informations diversifiées et fortement contextualisées grâce à des activités d'analyse organisationnelle, observation participante, suivi des projets, ainsi que la participation aux réunions internes et publiques liées aux plans, des discussions avec les élus, les fonctionnaires, les usagers et les citoyens associés aux plans. Cette étude de cas a été complétée par une série d'entretiens avec l'administration régionale, ainsi qu'avec des fonctionnaires dirigeants et des élus locaux dans les communes wallonnes engagées dans un PST³ (Pirson, 2015). L'auteur a aussi participé à une évaluation ex-post de la mise en œuvre des plans locaux de cohésion sociale (PCS) en Wallonie (Fallon, Italiano, Thiry et Das-sargues, 2013; Fallon et Feron, 2017) : réalisée à travers six études de cas et huit groupes de discussion avec les fonctionnaires responsables de ces plans, cette évaluation a permis de mettre en évidence, au plus près de la vision des acteurs concernés, élus, fonctionnaires et bénéficiaires, les initiatives innovantes en matière d'évaluation participative de l'impact des PCS.

Les sections suivantes présentent d'abord les processus de rationalisation que cette nouvelle instrumentation tend à renforcer au niveau des administrations locales, avant d'en détailler les dispositifs participatifs proposés à différents moments de l'action publique.

RATIONALISATION DE LA FONCTION PUBLIQUE LOCALE EN WALLONIE (BELGIQUE)

Au niveau régional wallon, les processus de gestion budgétaire imposent aux pouvoirs locaux une reddition détaillée des comptes et des indicateurs-clés sur l'évolution de la situation budgétaire et de la gestion des ressources humaines de l'entité. Par contre, les approches stratégiques et les pratiques d'évaluation des politiques sont encore limitées et peu institutionnalisées (Jacob, Varone et Genard, 2007). Mais depuis quelques années, les autorités locales, politiques et administratives, se voient confrontées à une multiplication d'injonctions pour une approche plus stratégique de l'action publique, encourageant le développement de démarches évaluatives. Ces démarches gestionnaires et stratégiques sont diffusées à travers des injonctions réglementaires, des programmes de financement des autorités régionales ainsi que des bureaux de consultants. Cette section présente un dispositif nouveau de gestion stratégique, emblématique des changements en cours, le « Programme stratégique transversal », et pose la question de sa compatibilité avec les transformations du statut des responsables administratifs locaux, pour dans un second temps mettre en évidence les éléments d'appropriation de ces instruments.

Le Programme stratégique transversal (PST)

Le PST fusionne en un document unique les ambitions de la déclaration politique communale⁴ (projets et programmation) avec un souci de planification et de gestion des ressources administratives et budgétaires, pour les six années de la législature. Il présente un diagnostic et les objectifs stratégiques, précisant le cadre opérationnel et le système de suivi des résultats, à la fois pour l'administration communale elle-même et pour les actions externes à réaliser.

Le ministre régional de tutelle ainsi que la fédération des pouvoirs locaux⁵, déclarent vouloir renforcer la professionnalisation de ces derniers : la rédaction et la mise en place d'un PST devraient encourager la planification des ressources, une gestion par projets, une évaluation continue des performances, ainsi qu'un dialogue informé avec les citoyens (UVCW, 2011).

Les projets repris dans le programme sont définis sur base de diagnostics locaux, entre élus et fonctionnaires, et leur suivi et évaluation relèvent de la responsabilité administrative. L'approche proposée pour le PST relève aussi d'une ambition managériale (Aucoin, 1990) : le volet politique est rédigé par le Collège des élus locaux, mais le programme d'actions repose sur un arbitrage entre les projets politiques et les ressources disponibles pour les réaliser. C'est aux fonctionnaires dirigeants de l'administration de rédiger le volet opérationnel à partir des objectifs stratégiques proposés par les élus : entre volet politique et volet administratif, les modes de rédaction sont différents, mais les contenus doivent s'articuler pour former à terme le « contrat d'objectifs » de l'administration.

Ces réformes modifient aussi la gestion interne de l'administration. La logique de gestion de projet se matérialise notamment par la déclinaison en mesures, actions et tâches à réaliser pour chacun des axes stratégiques de même

que par le développement de tableaux de bord pour organiser un suivi partagé. Dans un souci de reddition des comptes, le PST doit faire l'objet d'une évaluation après trois et six ans : des indicateurs de performance, au niveau des réalisations autant que des résultats, devraient aussi assurer une visibilité de l'action publique au niveau local.

La réforme par contractualisation de la fonction publique

En mettant en exergue la distinction entre volet politique (ou externe) et volet administratif (ou interne), l'approche PST est censée renforcer l'autonomie des responsables administratifs tout en améliorant la lisibilité de l'action de l'administration. Si l'autonomie gestionnaire du directeur général est renforcée, elle s'accompagne d'une obligation d'évaluation et de reddition des comptes. Ce nouveau dispositif doit être analysé dans le contexte de la réforme du statut des fonctionnaires dirigeants qui a été mise en place en même temps que le PST et qui a pour ambition de moderniser le fonctionnement des administrations locales en transformant la collaboration entre fonctionnaires dirigeants et politiques. Le directeur général négocie un contrat d'objectifs qui l'engage pour la durée de la législature et sur base duquel il sera évalué et qui intègre les objectifs politiques dans l'ensemble de ses missions ainsi que les moyens budgétaires et ressources humaines qui leur sont affectés (moyens et ressources).

Un élément marquant de la réforme est le projet d'introduction d'une procédure d'évaluation du directeur général, réalisée tous les trois ans, en référence à la description de fonction et à sa lettre de mission, sur la base de la qualité des actions menées en vue d'atteindre les objectifs fixés⁶. Un tel instrument contractuel introduit des inflexions proprement managériales dans un modèle de coopération politico-administratif fondé sur un partage de pouvoir caractéristique d'un système consociatif (Hood, 2000). La réforme impose aussi une nouvelle coordination des services pour renforcer la transversalité des politiques par la mise en place d'un comité de direction, comme organe administratif collégial et lieu d'échange interactif entre les responsables de l'administration. Pour améliorer l'intégration entre acteurs politiques et administratifs, la réforme propose la mise en place d'un comité de pilotage associant les élus locaux pour la rédaction et le suivi du PST.

Cette réforme des grades légaux, comme l'instauration du PST, marque un renforcement de la capacité discrétionnaire des agents communaux : les politiques définissent les projets et l'administration doit s'en saisir pour en assurer la réalisation (Aucoin, 1990). Il ne s'agit pas tant de rétablir la primauté des élus sur la bureaucratie, que d'affirmer la primauté des principes managériaux sur la bureaucratie, dans une logique d'efficacité et d'efficience, avec un rapprochement limité envers les usagers au nom de la qualité des services.

L'appropriation du PST par les communes pilote

En tant qu'outil de planification et de gestion stratégique, le PST s'est imposé assez facilement et de nombreuses communes pilotes se sont portées volontaires pour le mettre en œuvre dès 2013 :

En termes de gestion en tant que telle, c'est quand même un outil qui me paraît fondamental. Bon, je ne vais pas dire que c'est presque inventer l'eau chaude, mais c'est quand même étonnant qu'on n'ait pas eu ce type de démarche plus tôt. Comment gérer une commune si on ne sait pas les objectifs qui sont vraiment importants pour les décideurs, si on ne sait pas les moyens qu'on a alloués, et si on ne sait pas si on aura la capacité interne de gérer les choses? (un DG)

L'implémentation du PST a contribué à clarifier les rôles entre le pôle stratégique des élus et le pôle administratif :

Le PST aura permis de remettre un petit peu chacun à sa place, chacun dans son domaine; le politique continue à faire de la politique et prendre les décisions mais que l'exécution revienne bien à l'administration et pas de temps en temps, l'un comme l'autre, qui a tendance à changer un petit peu de route. (un DG)

En même temps, l'élaboration du PST permet de conscientiser les élus sur les contraintes propres à l'administration tandis que cette dernière prend conscience des objectifs à plus long terme définis par le politique. Cet aspect se révèle particulièrement bien dans les propos d'un DG :

Donc je pense que eux (le politique), ça leur permet de se rendre compte de notre réalité de terrain et que nous, on sait pas toujours tout faire, qu'il faut pouvoir se donner de temps en temps des délais. Et à l'inverse, je crois que l'administration s'est rendu compte aussi, ben que, le politique, il était là pour gérer. Il a des envies, il veut réaliser des choses. Mais je suis pas sûre que tout le monde savait ce qu'on avait envie de réaliser ici. Donc ça été un échange un peu, de dire : « voilà nos difficultés de terrain, voilà ce dont on a besoin et de l'autre côté « voilà ce qu'on voudrait ». Donc c'est vrai qu'on a essayé de faire matcher tout ça.

La réforme du statut des fonctionnaires dirigeants a renforcé leurs ressources en même temps que leurs responsabilités : l'interférence des autorités politiques dans leurs affaires en devient moins légitime. Il se forme un partenariat entre administration et élus : l'administration est invitée à co-construire le PST et ce faisant, elle contribue à conseiller le politique en amont des décisions, ce qui est nouveau pour certaines communes.

Le PST est bien un outil de pilotage à l'attention de l'administration elle-même. Auparavant, le seul document censé guider l'action de l'administration était le « Programme de politique générale », document de quelques pages qui n'établissait pas de ligne directrice claire pour l'administration : ce document était perçu comme une formalité sans signification concrète pour l'administration. A contrario, dans le cadre du PST, les différents agents sont appelés à recentrer leurs activités sur des objectifs à atteindre.

L'élaboration du PST semble à la fois séparer les fonctions politique et administrative et renforcer la qualité de la communication entre ces deux pôles : la préparation du PST est le fruit de concertations entre le collège, les fonctionnaires dirigeants et, dans certains cas, avec les responsables de services. Comme le met

en exergue un DG, la mise en place d'un PST met en évidence une articulation politique/administration où les services administratifs gagnent en visibilité :

Ça a changé aussi les relations entre le DG et l'administration puisque ce dialogue tripartite a eu lieu à ce moment-là. Le DG qui était toujours au courant je dirais des décisions politiques ou qui participe aux décisions politiques voit arriver maintenant un nouveau partenaire qui est son administration et qui est amenée beaucoup plus tôt dans la discussion autour d'objectifs politiques. Et donc entre le DG et l'administration, il y a aussi une relation qui était bilatérale qui devient tripartite et ça a aussi influencé par rapport au DG. Pas seulement par rapport au politique mais par rapport au DG puisque ici il y a plutôt la création d'une équipe plutôt que une relation hiérarchique verticale comme il existait auparavant. Cette relation transversale est une nouveauté. (un DG)

L'élaboration d'un PST consiste à fédérer les agents communaux autour du projet commun en leur conférant la responsabilité des actions qui les concernent. Les agents sont enjoins de s'approprier ces actions prioritaires et dans certaines communes, ce sont les agents des services eux-mêmes qui les ont rédigés comme le montre cet extrait :

Ce n'est pas le comité (de pilotage) qui va écrire les fiches puis les donner aux gens, je ne pense pas que c'est une bonne manière de fonctionner parce que les gens ne se sentent pas concernés. Là, ils (les agents) ont écrit les fiches eux-mêmes, ils ont défini le plan d'action, ils l'ont intégré, ils savent que ça fait partie de leur travail. (un DG)

Cet investissement des agents dans la préparation et le pilotage du PST favorise le décloisonnement des services, non seulement par une amélioration des communications interservices, mais aussi en définissant les projets de manière plus transversale en incluant le politique, la direction et les services administratifs :

La préparation du PST a quand même eu le mérite de réunir un peu plus de personnes que ce qu'on fait d'habitude. En fait, ça n'arrive pas souvent que les employés aient une vision globale, parce que souvent, c'est relativement cloisonné. Donc c'était pas mal de réunir tout le monde et voilà. (un DG)

Le PST impose de développer pour chaque action un indicateur spécifique, qualitatif ou quantitatif pour évaluer le degré d'atteinte des objectifs. Cette traduction d'objectifs en indicateurs de performance est difficile, que ce soit pour les élus ou pour l'administration, qui ne disposent pas des compétences nécessaires pour faire face à une nouvelle manière de penser la qualité de l'action publique. L'administration tend à se concentrer sur des indicateurs de réalisation, à savoir une analyse sur les prestations fournies relativement facile à interpréter.

Pour certains, on a essayé de mettre des indicateurs, mais ça c'était très difficile et donc ça, je ne pense pas qu'il y ait beaucoup de communes qui sont parvenues à mettre des indicateurs sur tout. Moi je pense que je vais plus fonctionner au niveau du taux de réalisation de mes fiches, ce sera mon indicateur. (un DG)

Les mêmes problèmes sont mis en avant pour ce qui relève de la contractualisation des fonctionnaires dirigeants. Par exemple, la rédaction de la lettre de mission, que doit transmettre le collège des élus au fonctionnaire dirigeant, est très souvent déléguée à ce dernier, ce qui rend floue la frontière entre pilotage stratégique (autorités politiques) et pilotage opérationnel (dirigeants administratifs).

La lettre de mission c'est le collège communal qui doit le remettre à son DG, pffffff ils disent « tu ne sais pas me la faire? Tu sais pas me la remettre et on corrigera? » Je dis : « oui mais demandez-la moi d'abord, pourquoi est-ce qu'ils ne me la demandent pas? » C'est pour ça que je vous disais la formation, la sensibilisation de la partie politique pour moi constitue un préalable qui n'a pas existé. (un DG)

Chaque commune ayant sa propre réalité locale, la mise en œuvre du PST ne s'est pas déroulée de la même manière ni avec la même intensité dans toutes les communes. Les entretiens menés auprès des communes pilotes qui ont rédigé un PST montrent que cette opération a permis des avancées telles que « l'orientation sur les objectifs, le décloisonnement des services, la planification des ressources budgétaires et humaines, la priorisation et l'attitude proactive » (Pirson, 2015, p. 28). Alors que les grandes communes avaient déjà mis en place une gestion plus transversale de l'action locale par la création d'organes collégiaux, les communes les plus petites ont rencontré davantage de difficultés, ne disposant souvent pas de l'expertise nécessaire pour réformer leurs méthodes de gestion ou n'arrivant pas à convaincre les services de la plus-value d'une gestion par projets.

Enfin quand même entre la petite commune où le bourgmestre est venu lui expliquer lui-même ce qu'il avait écrit lui-même dans son PST et la ville de La Louvière qui a désigné un directeur de la stratégie avec une cellule stratégie, on est dans deux mondes complètement différents hin... les DG dans les petites communes font un peu de tout que ça fait un peu peur pour eux quoi. Mais voilà j'ai pas vraiment d'avis clair et net sur le fait que ça devrait être obligatoire mais les petites communes hin je les plains en tout cas. (un DG)

La mise en œuvre de ces outils de planification et d'évaluation nécessite « une évolution non seulement des pratiques et outils de pilotage, mais également des habitudes prises et plus généralement de la culture politique en place, processus qui s'inscrit dans la durée » (Emery et Giauque, 2005, p. 130). En modifiant les rôles des dirigeants et des gestionnaires, un outil comme le PST induit un changement culturel profond ce qui peut ralentir la dynamique de réforme.

PARTICIPATION DES USAGERS POUR UNE DÉMOCRATISATION DE L'ACTION PUBLIQUE

L'approche managériale de l'administration publique défend l'idée d'une écoute active des usagers, invoquant l'idée que la définition des besoins publics et la programmation des actions ne peuvent résulter du seul processus électoral. Il s'agit

d'organiser une écoute directe et continue des « usagers », ainsi qu'une concertation avec les parties prenantes et avec la population.

Des espaces locaux de participation ouverts aux publics et censés créer, au sein d'un ordre politico-administratif anonyme et excluant, des espaces locaux de délibération et de formation d'une volonté collective se sont développés d'abord autour des projets de rénovation urbaine ou pour l'implication des associations et des riverains dans les projets d'aménagement (Talpin, 2013). Dans les années 70, la politique de la ville a favorisé les lieux participatifs au niveau des quartiers et créé des espaces de convivialité loin des pesanteurs bureaucratiques ou des tendances rationalisatrices. Les citoyens sont alors reconnus comme disposant de compétences spécifiques, d'une connaissance des problèmes locaux différente de celle des experts et qu'il importe de mobiliser dans l'action publique. Si l'élue conserve le monopole de la décision, la participation a aussi pour objectif d'accroître la proximité entre élus et citoyens, et d'intéresser la population aux questions locales dans un souci de lutte contre la « crise de la représentation ». Depuis lors se sont développées des techniques participatives internes aux services de l'administration, avec l'objectif à la fois d'améliorer les services publics et de restaurer le lien social (Chevallier, 2011). Cette tonalité participative reste marginale dans la démarche de programmation stratégique du PST, mais elle est mobilisée dans d'autres instruments sectoriels d'action publique locale tels que le Programme communal de développement durable (PCDR), ou le Plan de cohésion sociale (PCS). Ces deux plans, comme le PST, développent une démarche stratégique, avec des objectifs définis localement sur base d'un diagnostic initial, mais ils présentent en plus la caractéristique d'être gérés avec les habitants dans une logique participative. La question aujourd'hui posée est celle de leur contribution réelle à une plus-value démocratique.

Le Programme communal de développement durable (PCDR)

Le Programme communal de développement durable (PCDR) définit les lignes directrices d'un programme d'investissement financé par la Région wallonne (Godart et Deconinck, 2003), à travers « un processus participatif... (par lequel) les mandataires, la population, les associations, les acteurs économiques, sociaux, culturels et environnementaux élaborent et mettent en œuvre une stratégie pour leur territoire »⁷. La population doit être consultée dans un premier temps pour enrichir puis pour commenter les résultats du diagnostic socio-économique et environnemental initial et encadrer les évaluations des réalisations. Comme pour le PST, le programme doit proposer une vision stratégique, pour un horizon de dix ans cette fois, mais aussi définir une stratégie de développement énoncée sous forme d'objectifs et de projets, avec une planification temporelle et budgétaire⁸.

La procédure de définition du plan et des projets qui le constituent est particulièrement originale. Elle doit se faire en coopération avec les habitants réunis dans la Commission locale de développement rural (CLDR) composée de 10 à 30 membres représentatifs de la diversité locale (villages, professions, secteurs culturels, etc.) dans une logique de participation ascendante des habitants

chargés de proposer les axes de développement en mobilisant une vision à dix ans. Ces processus de participation avec la CLDR sont orchestrés quatre à six fois par an, tout au long du plan, avec le soutien d'une agence externe, spécialisée dans l'organisation de processus participatifs. Un élu local assure la présidence. L'administration communale assume le suivi des réalisations de projets, après approbation par l'assemblée communale.

Quand la dynamique de participation fonctionne bien, les citoyens arrivent à s'impliquer dans le développement des projets en évitant la politisation de la CLDR :

On a quand même au niveau participatif une série d'outils où ce sont des citoyens qui participent en fonction de leur intérêt, mais aussi en fonction des associations, des groupes de pression, etc. Cela roule assez bien parce que ce n'est pas du tout politisé, ou très peu. (entretien avec un échevin)

Des réalisations, vais-je dire, plus citoyennes. Mais, c'est – encore une fois – très très lourd comme mise en place. La consultation, ça nous a pris, faire le tour des villages, ça nous a pris deux ans. Et, pour peu de sensibilisation de la part de la population. (un DG).

Les procédures d'élaboration et de suivi d'un PCDR et d'un PST coïncident sur plusieurs points : réalisation d'un diagnostic préalable, définition et opérationnalisation des objectifs; planification temporelle formalisée sous forme de tableaux de bord précisant les indicateurs de réalisation et de résultats. Tous ces éléments concordent avec une vision davantage managériale de la gouvernance publique. Toutefois, l'originalité fondamentale du PCDR est de rendre obligatoire un dispositif complexe de coproduction avec un groupe de citoyens tout au long du plan et particulièrement lors du développement de celui-ci. Cette première étape peut durer près de deux années et impose un diagnostic très détaillé, une identification participative des objectifs stratégiques et une définition ex-ante détaillée de tous les projets. Pour ce faire, le décret impose le recours à des « quasi-professionnels » qui sont agréés en matière de participation, de communication et de gestion de projet. Dans le cas du PST, par contre, le recours à des experts extérieurs n'est pas imposé, ce qui devrait favoriser l'appropriation, par les agents communaux, de la logique d'approche stratégique, de gestion par projet et d'évaluation propre au PST. Les agents communaux sont donc censés disposer de multiples compétences (rédaction de diagnostics, évaluations, gestion de projets, etc.) pour administrer l'ensemble des étapes du processus, alors que, en pratique, ils disposent rarement de l'expertise nécessaire. Cette divergence entre PCDR et PST pose la question de l'encadrement à fournir aux fonctionnaires afin qu'ils puissent mener à bien une approche stratégique et des dispositifs participatifs.

Le Plan de cohésion sociale (PCS)

Le Plan de cohésion sociale (PCS) est un instrument régional de financement qui cofinance des programmes d'intervention orchestrés par les communes, dans une

logique de renforcement de l'égalité des chances et des conditions (Wayenberg, Reuchamps, Kravagna et Fallon, 2016).

Le plan est défini sur base d'un diagnostic qui identifie, en partenariat avec le secteur associatif local, les problèmes et besoins du territoire en matière d'emploi, de logement, de santé et de cohésion intergénérationnelle ou interculturelle. Les orientations stratégiques et les actions sont souvent proposées et gérées par des partenaires locaux et non par la commune, dans une logique de « réseau » territorial. Le plan initial doit être rédigé dans une logique stratégique et transversale, et construit puis évalué en partenariat, avec les associations et les bénéficiaires finaux. La coordination, la mise en œuvre et le compte rendu des actions sont assurés par un chef de projet qui est fonctionnaire communal, chargé de réaliser un état d'avancement annuel et une évaluation plus large la dernière année, en analysant, pour chaque action, l'adéquation entre les objectifs de l'action et les résultats obtenus. En outre, l'un des enjeux de l'évaluation des PCS réside dans la dimension dite « participative » de ceux-ci. La participation citoyenne est une notion centrale de la philosophie des PCS : elle est mise en avant aussi au moment de son évaluation, laquelle est organisée avec une partie des bénéficiaires et des partenaires associatifs, pour mesurer l'impact des actions sur la cohésion sociale et le bien-être tel qu'ils sont perçus au niveau local. Les procédures méthodologiques imposaient l'organisation de tables rondes, coordonnées par les chefs de projet, pour analyser avec les bénéficiaires dans quelle mesure les projets du PCS avaient modifié leurs indicateurs de bien-être (Charlier, Reginster, Ruyters et Vanden Dooren, 2014).

Cette évaluation participative avait pour objectif d'informer les responsables politiques avant la nouvelle programmation pour prendre en compte ces résultats dans le nouveau diagnostic (Fallon et coll., 2013). Si ce processus d'évaluation a été fort critiqué, la dynamique a néanmoins été appréciée comme une expérience positive par l'administration et les élus : « Ça a permis de crédibiliser les actions qui avaient été critiquées par le politique. Ça a aussi permis, en tant que nouvelle chef de projet, de légitimer mon travail » ; « [...] Mais elle nous a obligés à aborder des thématiques que nous n'aurions pas abordées autrement » ; « C'était lourd, mais c'était intéressant [...] Les attentes ont évolué – oui surtout au niveau des partenaires quand ils ont vu ce que les bénéficiaires ont répondu. Ils ont réfléchi... » (témoignages de chefs de projet – Fallon et coll., 2013). Cet exercice d'évaluation participative se distingue en effet de l'évaluation gestionnaire (celle qui est présente dans le PST) en poursuivant simultanément une finalité pratique et une finalité émancipatrice. La finalité pratique a pour but une meilleure utilisation des résultats de l'évaluation : en y participant, les parties prenantes sont censées s'approprier les résultats plus facilement. La finalité émancipatrice insiste davantage sur le caractère « endoformateur » de l'évaluation dont la notion d'apprentissage collectif est centrale (Plottu et Plottu, 2009). Le but principal de ce mode d'évaluation participative est donc de rompre avec l'approche traditionnelle souvent qualifiée d'« évaluation gestionnaire », dont le but premier, dans sa version instrumentale, vise principalement à aiguiller le décideur politique dans ses décisions. Les approches dites participatives ou pluralistes relèvent plutôt d'une

logique démocratique – dont le but ultime est l'accroissement de la possibilité de débat et d'intervention des citoyens dans le champ des politiques publiques – en tension avec la logique de l'efficacité décisionnelle (Corcuff, 1993) : l'intégration de la pluralité des systèmes de valeurs dont les acteurs sont porteurs interdit en effet la construction d'un jugement unique sur l'action et encourage l'intégration de l'évaluation au sein de la définition politique même, par des processus itératifs d'ajustements et de déplacements.

Ces deux exemples de dispositifs, PCDR et PCS, démontrent que la démarche participative au niveau local est possible, même pour organiser l'évaluation de l'action publique, mais qu'elle rencontre encore à la fois des tâtonnements et des oppositions. La mobilisation de ces principes participatifs génère des résistances tantôt du côté du Collège communal, qui n'admet pas toujours de voir limiter son pouvoir d'initiative et qui repousse alors les ingrédients de la démocratie directe et tantôt du côté de l'administration, qui ne dispose pas des compétences (dynamique de groupe, communication) pour organiser ces processus par ailleurs coûteux en temps et en ressources et qui tend à dénoncer l'adéquation de l'expertise des citoyens.

DISCUSSION

Au fil des décennies, les communes se sont transformées, avec une augmentation de la variété des secteurs d'intervention et une augmentation des processus de participation (Matagne, Radoux et Verjans, 2011). Les responsables administratifs et les élus doivent mobiliser des compétences de plus en plus complexes et techniques. Les communes wallonnes se voient encouragées à recourir à de nouveaux outils de gestion, sous l'impulsion de la Région wallonne : « La plupart des outils que j'ai mentionnés émanent en réalité de la Région wallonne. Le plus bel exemple maintenant est le PST. Le PST, c'est un magnifique outil. Chaque commune l'utilise à nouveau comme elle le souhaite » (un DG).

L'approche par les dispositifs d'action publique permet de mettre en évidence les évolutions des modes de gestion au niveau des pouvoirs locaux dans une logique socio-historique (Lacouette Fougère et Lascoumes, 2013). Notre analyse a été menée au plus près des pratiques, en tenant compte du contexte concret dans lequel les acteurs entretiennent des interactions et mobilisent des intermédiaires : observer un « instrument en action » permet de mettre en évidence non seulement ses effets spécifiques, mais aussi ses effets politiques et les relations de pouvoir qu'il génère. Les dispositifs imposés aux acteurs locaux pour la mise en œuvre de leurs politiques sont porteurs d'une certaine logique imposée par l'autorité régionale. Les procédures encadrant des projets d'investissement relèvent de la gestion classique, suivant un schéma administratif linéaire et séquentiel, sous l'impulsion politique, en mobilisant des processus de participation limités à l'information des riverains d'un projet dessiné par d'autres, sans mobiliser de dynamiques participatives ni de moments d'évaluation. De nouveaux instruments tendent à transformer ces pratiques et trois d'entre eux ont été présentés. Depuis plus de 25 ans, le PCDR

impose aux communes participantes une gestion différente : stratégique et participative, mais sans évaluation. Depuis quelques années, les dispositifs de gestion émergents (PCDR, PST, PCS) mobilisent tous une approche stratégique par projets, mais ils ne sont pas tous aussi exigeants que le PCS en matière d'évaluation ou que le PCDR pour la participation des citoyens.

Les trois dispositifs déploient des programmations stratégiques et opérationnelles sur base de diagnostics initiaux et imposent un suivi des réalisations :

- Depuis les années 1990, le PCDR a provoqué une ouverture remarquable à la participation des citoyens, depuis le diagnostic jusqu'à la mise en œuvre des projets, mais cette réforme ne concerne que des politiques ciblées;
- Les PCS imposent une évaluation de l'impact des actions menées au niveau local et cette évaluation doit être réalisée de façon participative, comme toutes les autres étapes du plan;
- L'instauration du PST provoque une révolution managériale interne de l'administration et redessine les processus de coopération entre les élus et les fonctionnaires; mais l'attention aux citoyens reste limitée à une logique de transparence renforcée de la reddition des comptes et d'une mise en visibilité des processus de gestion de projets.

Ces quelques exemples montrent que le processus de mise en œuvre du PST doit être replacé dans le cadre plus large des changements qui affectent actuellement le mode de gouvernance des administrations locales. En effet, les autorités communales se voient de plus en plus souvent contraintes (ou encouragées, suivant les cas) à développer leurs actions dans une logique stratégique. Cette approche stratégique de la gestion publique implique que les responsables locaux réalisent des diagnostics permettant d'identifier les caractéristiques locales pertinentes du problème à gérer, s'inscrivent dans une dynamique de gestion de projet consistant à mobiliser et à affecter les différentes ressources disponibles de façon adéquate, et procèdent à l'évaluation de la bonne gestion et de l'impact des actions réalisées avant d'établir la programmation suivante. Tous ces éléments témoignent d'une managérialisation accrue de la gouvernance locale en Région wallonne, transformation qui renouvelle fondamentalement la manière de construire une politique publique, en mobilisant une logique stratégique, adaptée au contexte local, et, dans une moindre mesure, participative.

L'analyse des dispositifs démontre une appropriation des nouveaux outils de management public et de mécanismes innovants de participation des citoyens et des usagers, en résonance avec les deux principes de transformation que soulignaient Pollitt et Bouckaert (2011) dans leur présentation de l'état neo-weberien : l'intégration des principes du managérialisme, pour plus d'efficacité et d'efficience, et une logique d'extension de la participation. Ces réformes favorisent-elles une plus-value démocratique? Les nouveaux instruments analysés ci-dessus organisent pour certains non seulement une extension de l'implication des citoyens,

mais aussi une délimitation plus stricte de l'autonomie politique des responsables élus et un renforcement de la discrétion administrative. Les interactions au sein du triangle entre administration – politique – citoyen sont donc reconfigurées.

Il a été montré ci-dessus que le PST a pour ambition d'assurer une nouvelle coordination entre administration publique et responsables politiques et que la réalisation de cet objectif a redéfini les responsabilités respectives des autorités administratives et des politiques élus, déléguant de nouvelles responsabilités aux fonctionnaires dirigeants tout en leur imposant une obligation plus stricte d'évaluation et de reddition des comptes. Quelques années plus tôt, une première réforme du Code de la Démocratie Locale et de la Décentralisation (CDLD) s'attellait à relégitimer le lien entre le bourgmestre et la population par un processus électif qui atténue les effets de tête de liste et donc le contrôle des partis (Matagne et coll., 2011) mais la réforme récente du Code renforce l'autorité des responsables administratifs non seulement sur la gestion interne de l'organisation, mais aussi sur la production politique. Ce dernier processus qui relève plutôt d'une ambition de rationalisation managériale ne laisse pas de poser question au niveau des politiques : les élus locaux sont généralement en communication assez étroite avec les usagers et les bénéficiaires des politiques qu'ils ont eux-mêmes décidées et qui sont mises en œuvre par une administration elle aussi locale, mais moins habituée aux logiques délibératives avec les usagers. Dans certaines communes, la différenciation des rôles politiques et administratifs qu'essaie de favoriser le PST rencontre un réel frein du Collège des élus qui tient à maintenir ses positions de pouvoir aussi au sein de l'administration.

Par exemple, le PST est censé favoriser la compréhension de l'action politique par les citoyens. Les communes publient leur PST en ligne, tantôt intégralement, tantôt dans une version résumée des objectifs stratégiques. Tout en améliorant la visibilité de l'action publique, les élus communaux sont conscients qu'ils s'exposent davantage aux critiques de leurs électeurs, dans une logique de reddition des comptes, mais aussi de l'opposition. Celle-ci dispose, lors de l'évaluation du PST, d'un tableau de bord reprenant les réalisations de la législature, au regard des objectifs déclarés :

C'est aussi se mettre à nu pour un politique parce que c'est annoncer clairement les projets qu'il veut mettre en œuvre et donc c'est aussi une base sur laquelle il pourra être évalué et donc parfois certains politiques préfèrent avancer sans trop se dévoiler pour ne pas être face à un échec. (un DG)

Les nouvelles exigences de participation imposées par les autorités régionales entraînent des coûts (en temps ou en ressources humaines) auxquels l'administration ne peut pas toujours faire face. Chevallier (2011) souligne l'importance du renforcement du lien entre administration et citoyen, particulièrement au niveau local : le renforcement des dynamiques participatives pourrait favoriser la mise en place d'une démocratie renouvée où le citoyen se présente moins « assujéti » à l'administration et davantage porteur de droits exigibles, réels et concrets, dont il demande de plus en plus le respect dans sa singularité. Ces processus favorisent

un nouveau regard de l'administration sur les citoyens et inversement. Mais la question reste posée pour les fonctionnaires, d'une possible intégration dans la rationalité gestionnaire des structures politico-administratives de ces nouveaux processus de coopération avec les citoyens locaux, que ce soit dans un souci de meilleure gestion, quand la participation des citoyens renforce la qualité de l'action publique, ou pour gérer l'affrontement avec des options divergentes lors des conflits d'implantations. Ils doivent de plus en plus intégrer dans les procédures de décision un (ou plusieurs) moment délibératif associant les publics concernés en amont de celle-ci pour faire remonter vers les décideurs des informations quant aux demandes des usagers.

L'administration se voit donc enjoindre d'améliorer son accessibilité, mais aussi sa transparence, tantôt en interne, envers les institutions de l'État lui-même, tantôt à l'externe, vers les citoyens. Les autorités régionales, qui subsidient en grande partie les programmes analysés et qui exercent une certaine tutelle sur les administrations locales, imposent de plus en plus la définition d'objectifs mesurables et la mise en place de tableaux de bord avec des indicateurs d'avancement, de réalisation et de résultats, pour accompagner la mise en place des plans stratégiques, multipliant les possibilités de contrôle des représentants sur l'administration. Cette augmentation des contraintes en matière de transparence et d'évaluation de la performance pourrait apporter elle aussi une certaine plus-value démocratique, mais dans un sens différent (Ficet, 2017). Ces outils généralisent la circulation des informations concernant les activités de l'administration pour éventuellement favoriser leur mise en débat au sein de la société locale. Cette dimension externe, voire démocratique, de la transparence renvoie à la capacité des citoyens à influencer les processus gouvernementaux grâce à un accès facilité à l'information gouvernementale et aux instances décisionnelles, pour in fine améliorer la qualité de la gestion publique, leur confiance dans cette dernière et la légitimité de l'État. Mais, ces communications favorisent d'abord les mécanismes de transparence technique pour montrer au citoyen la qualité et l'efficience de la gestion publique dans une logique de redevabilité administrative. Une telle logique de transparence met l'accent d'abord sur les résultats et les performances de la gestion publique, mais sert si peu les intérêts directs des citoyens qu'il pourrait, paradoxalement, contribuer à la détérioration de la confiance des citoyens envers l'État (O'Neill, 2002).

Certains analystes en viennent à mettre en doute le potentiel démocratique de ces pratiques.

Les moyens mobilisés au niveau local permettent non seulement des enquêtes et des consultations (dont les classiques enquêtes de voisinage dans le cadre des permis d'environnement ou d'urbanisme), mais aussi des processus de co-construction : budgets participatifs, jurys citoyens, etc. Par exemple, la procédure participative du PCDR, très formalisée, relève bien de cette vision de co-construction d'un plan d'action – dont les représentants élus portent toujours la responsabilité formelle. Le PCS aussi associe les acteurs locaux tout au long du processus décisionnel local, instaurant une véritable approche ascendante, et ce

jusqu'à associer les bénéficiaires à l'évaluation de l'impact de ses réalisations. Dans le cadre du PST, par contre, le choix de faire appel ou non à des acteurs locaux reste aujourd'hui une décision politique, le processus participatif n'étant pas formellement obligatoire.

Les responsables administratifs ne manquent pas de souligner les coûts de la participation, le supplément de complexité et le ralentissement du processus de décision (Moynihan, 2003) et les difficultés à convaincre les acteurs responsables du secteur de la nécessité d'ouvrir celui-ci, quitte à menacer leurs propres positions. Il est bon de souligner un facteur souvent sous-estimé : les résistances locales de la population et la non-participation (Jacquet, 2017). Les populations ne sont pas toujours désireuses de s'engager dans des processus de délibération relatifs à l'action publique et le fonctionnement concret des instruments participatifs tend à privilégier certains groupes qui disposent de capital culturel ou d'une expérience militante. Moins de 1 % de la population participe à ces dispositifs et la plupart sont déjà investis dans d'autres espaces, associatifs, partisans ou scolaires (Talpin, 2013); ils sont généralement peu représentatifs de la population, par exemple avec une sous-représentation de jeunes ou de classes populaires. Ces dispositifs participatifs, s'ils renforcent la prise en compte d'un plus large pluralisme dans l'action publique, peinent à redistribuer le pouvoir de décision au niveau local. L'élus tend à se présenter comme l'unique détenteur de l'intérêt général, ce qui ne favorise pas la co-décision.

CONCLUSION

Les réformes de la gestion politico-administrative des communes wallonnes transforment profondément les modes de coopération entre les responsables locaux politiques et administratifs et avec les citoyens, mais les effets ne peuvent pas être analysés en utilisant directement les cadres d'analyse disponibles au niveau national, parce que les dynamiques interactionnelles entre les acteurs peuvent se révéler beaucoup plus intenses et complexes au niveau local. C'est en analysant les dispositifs d'action publique locaux d'une façon contextualisée et concrète, au plus près du terrain, qu'il est possible de mettre en évidence les rapports de pouvoir qu'ils maintiennent ou modifient.

Si la plupart des dispositifs étudiés démontrent une ambition de managérialisation de la gestion locale pour une meilleure efficacité de celle-ci, la question d'un potentiel renforcement démocratique du régime de gestion publique local à travers ces réformes demande d'étudier au plus près ces trois qualités emblématiques des démocraties libérales contemporaines : participation des citoyens, contrôle de l'autorité politique et de l'arbitraire de l'administration.

Si la participation des citoyens a été sollicitée depuis des dizaines d'années au cours d'enquêtes publiques accompagnant les projets d'aménagement et d'urbanisme, les dispositifs aujourd'hui multiplient les procédures imposant une co-construction des plans et des projets en partenariats avec des citoyens et/ou des associations locales, pour favoriser à la fois la qualité des actions et leur

appropriation par les utilisateurs. Mais leur mise en œuvre effective reste encore relativement difficile. D'une part, ces dispositifs participatifs exigent de nouvelles compétences et des ressources au sein de l'administration, dans une logique d'intervention sociale plus que technique, et ils peuvent alourdir les procédures parfois au détriment d'une certaine efficacité : les réticences peuvent être fortes, contestant la logique même de l'ouverture pour assurer le maintien d'une gestion légale rationnelle orchestrée de façon purement bureaucratique. D'autre part, une très faible proportion seulement des citoyens choisit d'y participer, et ce, de façon intermittente (Jacquet, 2017). Et c'est au niveau infrapolitique, celui des pratiques individuelles concrètes des citoyens, au niveau des tensions organisationnelles dans les administrations, que les modalités de ces réticences se donnent à voir, alors que les principes mobilisés par ces dispositifs ne sont pas contestés (LeBourhis et Lascoumes, 2014).

Les dispositifs de managérialisation contribuent à renforcer la lisibilité de l'action publique, ce qui devrait permettre des débats mieux informés. L'administration cherche davantage à communiquer sur ses activités et ses résultats, dans une logique de reddition des comptes qui pourrait avoir l'effet de renforcer le contrôle politique. De nouveaux dispositifs participatifs lui sont imposés, pour associer à son action un public qu'il peine à constituer. Les réformes du statut des fonctionnaires dirigeants ont pour effet de renforcer leur prise sur leur propre administration ainsi qu'une autonomie plus large pour leurs orientations opérationnelles, dans une logique d'efficacité et d'efficence, qui pourrait augmenter leur pouvoir discrétionnaire : celui-ci reste néanmoins sous contrôle politique à travers les nouvelles obligations d'évaluation et de contrôle des responsables politiques et la logique contractuelle plutôt que hiérarchique qui habille aujourd'hui à nouveaux frais la relation entre responsables politiques et administratifs au niveau local.

L'analyse de l'évolution récente de l'instrumentation des dispositifs de gestion de l'action publique au niveau local montre l'émergence d'un régime associant un nombre plus important d'acteurs et une redéfinition des processus de communication entre citoyens, administrations et représentants, signe d'un rééquilibrage des rapports de pouvoir. Ces transformations s'organisent aujourd'hui à petits pas, mais non sans résistance, et pourtant sans bruit, entre autres parce que ce niveau de l'action publique intéresse moins les analystes politiques.

NOTES

- 1 Cet article a fait l'objet d'une présentation orale au cours du CoSPoF 2017 – Congrès annuel de la Société québécoise de science politique 2017, « Flux et Frontières. Réponses politiques et identitaires – ST : Démocratie et politiques publiques »
- 2 Il s'agit d'une commune rurale de petite taille (4500 habitants; administration locale occupant 50 équivalents temps plein) dont le niveau socio-économique est très légèrement supérieur à la moyenne régionale; située à 50 km de deux villes majeures, elle présente une démographie croissante.

- 3 Pour la période 2013–2018, la mise en place d'un Programme stratégique transversal se faisait sur base volontaire, mais avec le soutien de la Région wallonne et de l'UVCW (Boverie, Maître, et Van Driessche, 2013a) : le dispositif sera rendu obligatoire pour la mandature 2019–2014, suite à la modification du Code de la Démocratie Locale et de la Décentralisation (CDLD).
- 4 Équivalent du niveau municipal.
- 5 Union des Villes et Communes de Wallonie ou UVCW.
- 6 Si la loi (CDLD) prévoit cette évaluation, aucune commune ne l'a mise en œuvre à ce jour.
- 7 Art. 2, § 1^{er}, du décret wallon du 11 avril 2014 relatif au développement rural (*Moniteur belge*, 2 juin 2014).
- 8 Art. 13 du décret wallon du 11 avril 2014 relatif au développement rural.

BIBLIOGRAPHIE

- Aucoin, P. (1990). Administrative reform in public management: Paradigms, principles, paradoxes and pendulums. *Governance*, 3(2), 115–137. <https://doi.org/10.1111/j.1468-0491.1990.tb00111.x>
- Boverie, M., Maître, A. et Van Driessche, L. (2013a). Gouvernance – Fiche 2 – Le PST : Programme stratégique transversal communal. Dans *Focus sur la commune – 161 fiches pour une bonne gestion communale*. Namur: Union des Villes et Communes de Wallonie.
- Charlier, J., Reginster, I., Ruyters, C. et Vanden Dooren, L. (2014). *Indicateurs complémentaires au PIB : L'indice des conditions de bien-être (ICBE)*. Éditeur IWEPS. Récupéré de <https://www.iweps.be/>
- Chevallier, J. (2011). De l'administration démocratique à la démocratie administrative. *Revue française d'administration publique*, 1(137–138), 217–227. <https://doi.org/10.3917/rfap.137.0217>
- Chouinard, J. A. et Milley, P. (2016). From new public management to new political governance: Implications for evaluation. *CJPE*, 30(1), 1–22. <https://doi.org/10.3138/cjpe.30.1.1>
- Corcuff, P. (1993). Un O.V.N.I. dans le paysage français ? Éléments de réflexion sur l'évaluation des politiques publiques en France. *Politix*, 6(24), 190–209. <https://doi.org/10.3406/polix.1993.1596>
- Dahler-Larsen, P. (2012). *The evaluation society*. Stanford, CA: Stanford University Press.
- Emery, Y. et Giauque, D. (2005). *Paradoxes de la gestion publique*. Paris: L'Harmattan.
- Fallon, C., Italiano, P., Thiry, A. et Dassargues, A. (2013). *Rapport final : Évaluation des processus de mise en œuvre des Plans de Cohésion Sociale*. Éditeur Iweps – DiCS.
- Fallon, C., Hansoul, F. et Piron, D. (2016). Programme stratégique transversal (PST) et gouvernance locale en Wallonie. *Courrier hebdomadaire du CRISP*, 14(2299), 5–52. <https://doi.org/10.3917/cris.2299.0005>
- Fallon, C. et Feron, P. (2017). Quantification des politiques versus démocratisation de l'évaluation : Étude des plans de cohésion sociale en Wallonie. Dans M. Mespoulet (éds.), *Usages des chiffres dans l'action publique territoriale*. Rennes, France: PUR.

- Fallon, C. et Ficet, J. (éds.). (2017). *Crise de la démocratie et nouveau management public : discours, pratiques, instruments*. Louvain-la-Neuve, Belgique: Academia – L'Harmattan – Coll. ABSP.
- Ficet, J. (2017). Nouveau management public et démocratie. Une approche par les instruments. Dans C. Fallon et J. Ficet (éds.), *Crise de la démocratie et nouveau management public : discours, pratiques, instruments*. Louvain-la-Neuve, Belgique: Academia – L'Harmattan – Coll. ABSP.
- Godart, M-F. et Deconinck, M. (2003). Développement territorial en milieu rural : quelques exemples en région Wallonne. *Revue d'Économie Régionale & Urbaine*, 5(décembre), 909–924. <https://doi.org/10.3917/reru.035.0909>
- Hood, C. (2000). Relationships between Ministers/politicians and public servants: Public services bargains old and new. Dans B. G. Peters et D. Savoie (éds.), *Governance in the twenty-first century: Revitalizing the public service* (pp. 178–204). Montréal, QC: McGill-Queen's University Press.
- Jacob, S., Varone, F. et Genard, J. L. (éds.). (2007). *L'évaluation des politiques au niveau régional*. Bruxelles: Peter Lang, Coll. Action publique.
- Jacquet, V. (2017). *Délibérer ou ne pas délibérer ? La (non-)participation dans les dispositifs délibératifs tirés au sort*. Thèse défendue à l'UCL en vue de l'obtention du grade de Docteur en sciences politiques et sociales.
- Lacouette Fougère, C. et Lascoumes, P. (2013). L'évaluation : un marronnier de l'action gouvernementale ? *Revue française d'administration publique*, 4(148), 859–875. <https://doi.org/10.3917/rfap.148.0859>
- Lascoumes, P. et Le Galès, P. (Eds.). (2004). *Gouverner par les instruments*. Paris: Presses de Sciences Po.
- LeBourhis, J. P. et Lascoumes, P. (2014). Les résistances aux instruments de gouvernement. Essai d'inventaire et de typologie des pratiques. Dans C. Halpern et coll. (éds.), *L'instrumentation de l'action publique : Controverses, résistances, effets* (pp. 493–520). Presse de Sciences Po.
- Matagne, G., Radoux, E. et Verjans, P. (2011). La composition du collège communal après la réforme du Code wallon de la démocratie locale. *Courrier hebdomadaire du CRISP*, 9(2094), 5–35. <https://doi.org/10.3917/cris.2094.0005>
- Moynihan, D. P. (2003). Normative and instrumental perspectives on public participation: Citizen summits in Washington, D.C. *The American Review of Public Administration*, 33(2), 164–188. <https://doi.org/10.1177/0275074003251379>
- O'Neill, O. (2002). *A question of trust*. Cambridge: Cambridge University Press.
- Pirson, L. (2015). *Implémentation du projet pilote PST dans les communes wallonnes selon le paradigme de la Nouvelle Gestion Publique : critiques, limites et paradoxes*. Master Thesis en Science politique : administration publique (U.Liege), sous la direction de Fallon C.
- Plottu, B. et Plottu, E. (2009). Contraintes et vertus de l'évaluation participative. *Revue française de gestion*, 35(192), 31–44. <https://doi.org/10.3166/rfg.192.31-44>
- Pollitt, C. et Bouckaert, G. (2011). *Public management reform, a comparative analysis* (pp. 75–125). Oxford: Oxford University Press.

- Talpin, J. (2013). La démocratie participative marginalisée par le pouvoir local. *Savoir/Agir*, 3(25), 23–30. <https://doi.org/10.3917/sava.025.0023>
- UVCW - Union des Villes et Communes de Wallonie asbl. (2011). Le programme stratégique transversal communal (PST) Une démarche stratégique pour les villes et communes. *Mouvement Communal*, 863(hors-série), 2011.
- Wayenberg, E., Reuchamps, M., Kravagna, M. et Fallon, C. (2016). Policy analysis at local level: steering from below? Dans M. Brans et D. Aubin (éds.), *Policy analysis in belgium*. The International Library of Policy Analysis, Book Series.
- Weiss, C. H. (1999). The interface between evaluation and public policy. *Evaluation*, 5(4), 468–486. <https://doi.org/10.1177/135638909900500408>

PRÉSENTATION D'AUTEUR

Catherine Fallon est ingénieure civile et docteure en sciences politiques et sociales. Elle est chargée de cours au département de science politique de l'université de Liège, où elle dirige l'unité de recherche SPIRAL. Ses travaux actuels portent sur l'analyse et l'évaluation des politiques publiques et l'innovation dans l'administration publique, avec une attention particulière dans le domaine des politiques scientifiques en Belgique, la gestion des risques et la planification d'urgence. Elle codirige le Certificat interuniversitaire en évaluation des politiques publiques.

“Deliverology” and Evaluation: A Tale of Two Worlds

Lisa Birch and Steve Jacob

Université Laval

Abstract: *In recent years, the new political governance, a partisan model that contributes to a permanent campaign, gained ground in public organizations. In this new context, “deliverology” is portrayed as an innovative method to help governments implement new policies and deliver on election promises. This article presents the similarities and differences that exist between “deliverology” and evaluation. Is deliverology really something new or is it another case of old wine in a new bottle? Is deliverology a substitute for or, instead, a complement to institutionalized evaluation? To what extent does new political governance (exemplified by deliverology and performance measurement) undermine evidence-based decision making? What is the value-added of deliverology? These questions are addressed through a critical reflection on deliverology and its value-added in Canada, where evaluation became institutionalized in many departments and agencies under the influence of results-based management, promoted by the advocates of new public management over four decades.*

Keywords: *deliverology, election promises, evaluation, new political governance, results-based management*

Résumé : *Depuis quelques années, la nouvelle gouvernance politique, un modèle de gestion partisane qui contribue à une campagne permanente, gagne en faveur dans les organisations publiques. Dans ce nouveau contexte, la « résultologie » est présentée comme une méthode innovatrice de mise en œuvre des politiques gouvernementales et de suivi de la réalisation des promesses électorales. Cet article présente les points communs et les différences qui existent entre la résultologie et l'évaluation. La résultologie est-elle vraiment quelque chose de nouveau ou est-ce seulement une autre manière de réinventer la roue? La résultologie est-elle un substitut ou plutôt un complément à l'évaluation qui s'est progressivement institutionnalisée dans l'appareil d'État? Dans quelle mesure la nouvelle gouvernance politique (incarnée par la résultologie et la gestion du rendement) vient-elle miner la prise de décisions éclairées par des données probantes? Quelle est la valeur ajoutée de la résultologie? Pour répondre à ces questions, le présent article propose une réflexion critique sur la résultologie et sa valeur ajoutée au Canada, où l'évaluation s'est institutionnalisée dans les ministères et les agences sous l'influence de la tendance de gestion axée sur*

Corresponding author: Lisa Birch, Département de Science politique, Université Laval, Pavillon Charles-De Koninck, local 3449, 1030, avenue des Sciences-Humaines, Québec, QC G1V 0A6; lisa-maureen.birch@capp.ulaval.ca

les résultats, dont les partisans de la nouvelle gestion publique font la promotion depuis quelque 40 ans.

Mots clé : *résultologie, promesses électorales, évaluation, nouvelle gouvernance politique, gestion axée sur les résultats*

Throughout his latest book, entitled “What is Government Good at?: A Canadian Answer,” [Donald Savoie \(2015\)](#), a well-known Canadian political scientist, is highly critical of the inverted 90 percent–10 percent culture¹ with a strong, structural fault line between ideas and implementation that has arisen with the advent of the permanent election campaign. He argues that 90% of policy resources should be dedicated to implementation and only 10% to ideas. The current inversion of these proportions has accentuated executive-level centralization and partisan dynamics in Canadian public administration in the context of an increasingly complex, globalized policy environment with rising pressures for higher transparency standards and ever more demanding media scrutiny. [Chouinard and Milley \(2015\)](#) raise concerns about how this twenty-first-century trend, coined the “new political governance” by [Aucoin \(2012\)](#), undermines the evaluation function and evidence-based approaches in policy formulation, generating two paradoxes: (1) the paradox of use, which shifts the focus toward symbolic utilization for technical tracking, as opposed to policy learning and social betterment; and (2) the paradox of accountability, which redirects attention toward democratic accountability and away from accountability for public management, leading to the politicization of public administration and an erosion of public-sector norms of neutrality, independence, and professionalism.

In *How to Run a Government so that Citizens Benefit and Taxpayers Don't Go Crazy*, Sir [Michael Barber \(2016\)](#), the founder of “deliverology,” criticizes what he calls “government by spasm,” which is caught up in present-focused crisis management, producing sound bites and announcements in anticipation of and reaction to “noise” from the conventional and social media agendas. He advocates instead for “government by routines,” which is focused on managing and driving its own delivery agenda, with regular reporting on progress and success stories to the media and the public.

Recourse to Barber’s “deliverology” as a tool for new political governance began with the Tony Blair government in the United Kingdom (1997–2007) and spread to Canada through the Dalton McGuinty government (Ontario, 2003–13) and now the Justin Trudeau government (Canada, 2015–present). However, there is a dearth of evidence on deliverology’s effectiveness as a “way to run a government” and little analysis of how it is differentiated from results- or performance-based management or from concomitant evaluation or monitoring ([Richards, 2018](#); [Richards, Gallo, & Kronick, 2017](#); [Schacter, 2016](#)). In the Canadian context, where the state progressively institutionalized evaluation and results-based approaches over four decades, one may wonder about the “value-added” of Barber’s deliverology and, as [Schacter \(2016\)](#) puts it, what problem deliverology is purporting to solve.

Although [Savoie \(2015\)](#) and [Barber \(2016\)](#) share a similar diagnosis of modern governance challenges and the need to focus on implementation (see note 1), they do not reference each other's work or even the same literature on governance. Furthermore, where Savoie criticizes new political governance and private-sector-style performance management as harmful for implementation and service delivery in the public sector, deliverology embraces both the governing party's reform agenda and private-sector-inspired performance management for optimal delivery, thereby becoming a tool for new political governance. These observations inspire the image of two worlds, each with its own referential frame and its own conclusions, which share a vision that government should realign to deliver whatever it is good at doing for the greater good of citizens. The small overlap between the two worlds remains a concern for improved decision making regarding policy implementation and for better policy outcomes. One domain is the "deliverologist's" world and the other is the world of the evaluators, public administration scholars, and political scientists.

Through an overview of central ideas and criticisms of Barber's deliverology, we situate deliverology relative to *ex ante*, concomitant, and *ex post* evaluation as well as results-based management. We then explore the controversial implementation of deliverology in Canada. Despite their common concerns for improved policy implementation, deliverology and evaluation appear largely as two worlds with different roots and evidence standards and with some overlapping zones, notably the use of performance indicators in evaluation and the aspiration for social betterment through effective policy delivery and meaningful public-sector reform. However, there are vast zones where these worlds diverge discursively, methodologically, and philosophically.

WHAT IS BARBER'S DELIVEROLOGY?

Deliverology is a trademark created and owned by Delivery Associates, Michael Barber's consultancy firm ([Delivery Associates, n.d.a](#)). Although Barber initially spoke of an emerging science of delivery, Delivery Associates market it as follows:

Deliverology® is our system for helping governments deliver meaningful results that will last. It is both a science and an art. The science is the routine of setting a target and then using data, technology, planning, monitoring and problem solving to achieve it. The art is the way you do it and how you behave—it needs focus, urgency, ambition, honesty and humility. ([Delivery Associates, n.d.b](#))

Deliverology is a global business that promotes a model of how to run a government drawn from business management and its founder's practical experience in education and advising governments on delivering reforms. Barber relentlessly transformed his original practitioner experience acquired in the Blair years as head of the Prime Minister's Delivery Unit into a tool for global career mobility through partnerships and networks involving high-profile management consultant firms such as McKinsey & Company with the McKinsey Global Institute, and

the Boston Consulting Group with its Centre for Public Impact ([Brown, 2015](#); see also [Centre for Public Impact, 2016, 2017, n.d.](#)), and now his own Delivery Associates. This network includes international organizations like the OECD, the World Bank, the World Economic Forum (see [Global Agenda Council on Education, 2016](#), for an example), and non-profit or philanthropic think tanks such as the Centre for American Progress and the Brookings Institute, as well as The Economist Intelligence Unit, the Institute for Public Policy Research, the Institute for Government, the Social Market Foundation, and NESTA, all in the United Kingdom. Deliverology is thus an example of a business in the lucrative, public-sector management reforms industry (see [Manning, 2015](#), about this industry, and [Manning & Watkins, 2013](#), for a comparison of deliverology to other approaches).

The main sources of inspiration for Barber's model of delivery, as shown by his bibliographies, span an eclectic mix, including (1) his own experiences as recounted in his publications, (2) historical accounts of country success, fixing failures, and trends in Western history, with works by Fukuyama, (3) biographies or autobiographies of leaders such as Winston Churchill, Theodore Roosevelt, Margaret Thatcher, and Tony Blair, (4) various publications by governments, consultancy groups, think tanks, and international organizations, (5) popular literature on great or successful companies, (6) accounts of delivery-unit successes in other countries such as Chile and Malaysia, and (7) books on public and private strategies (see reference lists in [Barber, 2008, 2013, 2016, 2017](#); [Barber, Donnelly, & Rizvi, 2013](#); [Barber, Moffit, & Kihn, 2011a, 2011b](#)). The preference for grey literature as opposed to peer-reviewed academic publications is summarized in the following statement:

synthesis is now provided by organisations outside universities—thinktanks, public agencies or consultancies, or by those who translate the synthesis into action.... [T]he most influential reports globally in the past five years have come from the Organisation for Economic Co-operation and Development (OECD), McKinsey or, recently, the Economist Intelligence Unit. ([Barber et al., 2013](#), p. 17)

The academic literature in evaluation, public administration,² and political science is remarkably absent from Barber's reference lists. Conversely, references to "deliverology" *per se* are rare in evaluation and policy studies, where there is a vast literature grounded in scientific inquiry and practitioners' experiences after four decades of institutionalization and professionalization of evaluation in Canadian public administration ([Canada, Treasury Board Secretariat, 2013–2014](#); [Cummings & Gauthier, 2016](#); [Dobell & Zussman, 2018](#); [Jacob, 2006](#)). Thus, the first divergence between evaluation and deliverology is rooted in their respective sources of knowledge and inspiration.

DIVERGENT EVIDENCE STANDARDS

Flowing from this divergence is another one concerning evidence standards for policymakers. At first glance, deliverology seems at odds with discourses

on bringing scientific evidence back into policy. However, [Cairney and Oliver \(2017\)](#) remind us that scientists compete with other influential actors to access policymaking audiences. They argue that these audiences face two challenges. First, policymakers operate in complex, power-sharing, multi-level governance environments. Second, given this complexity and its inherent ambiguities, policy-makers act under conditions of bounded rationality and thus accept variations in evidence standards and hierarchies. Policymaking audiences can opt for evidence from implementation science, with its emphasis on gathering data on effectiveness, with attentiveness to evidence hierarchies that value systematic reviews and randomized controlled trials (RCTs), on scaling up best practices through replication, and on managing dosages of the "active ingredient" for change. They can also include other social science research as well as internal and external evaluations of policies and their implementation and outcomes ([Cantelli, Jacob, Genard, & de Visscher, 2006](#); [Nilsen, Ståhl, Roback, & Cairney, 2013](#)). They can choose evidence presented through storytelling approaches about indicators of effectiveness from practitioners and service users, narrations of experiences in case studies, and descriptions of key principles to be prioritized for service delivery (for an example, see *Results and Delivery—Lessons From Around the World*. [Canada School of Public Service, 2016a](#); see also [Delivery Associates, n.d.b](#)). In Canada, at the outset of his first mandate, Justin Trudeau, as Prime Minister of Canada, turned to Barber with his storytelling approach and more supple evidence hierarchy (see [Dobell & Zussman, 2018](#); [Puttick & Ludlow, 2013](#); [Shepherd, 2018](#)) instead of contracting to experts in public policy and public administration for guidance in implementing, rather than delivering, the Liberal Party of Canada's platform. It is important to note that, for Barber, evidence that a policy is working lies in achieving a set of chosen targets set by the executive level of government, whereas for evaluators and scholars, the aspirational standards are set in light of the policy goals and require the production of valid and reliable data using scientific methods to determine the policy's relevance, effectiveness, efficiency, impact, and sustainability in achieving positive intended outcomes.

IS DELIVEROLOGY A MODEL OF POLICY CHANGE OR OF PLEDGE FULFILLMENT?

Let us now turn to the core elements of deliverology as a model of policy change and implementation and as a tool for reforming public-sector management. This model tells policymakers at the top to begin by answering five simple questions: (1) What are you trying to do? (2) How are you trying to do it? (3) How, at any given moment, will you know that you are on track? (4) If you are not on track, what are you going to do about it?, and (5) How can we help?, and to establish a routinized system requiring civil servants to report upwards on these same questions ([Barber, 2008](#); [Canada School of Public Service, 2016a](#); [Murekeyisoni, 2017](#)). The governing party answers these questions in reference to its election platform, giving a partisan orientation to the policy goals, programs, and reforms

to be implemented by civil servants. The implicit, normative assumption is that the governing party's policy choices and reforms are inherently correct and fully legitimate, thus potentially precluding *ex ante* assessments of their relevance, logic models, and evidence base.

For the next step in implementation, Barber advises politicians that they must achieve the four components of deliverology in order as follows: (1) develop a foundation for delivery, (2) understand the delivery challenge, (3) plan for delivery, and (4) drive delivery (Barber et al., 2011a, p. 237). The first step involves reviewing the current state of delivery and creating a specialized results and delivery unit at the apex of power that leads delivery. Located deep inside the executive level of government, the delivery unit is far above Savoie's (2015) fault line in the current 90% range dedicated to policy ideas and communications. Although it does not invert the problematic "90/10" split, it does bring a strong mandate to pursue implementation relentlessly down the delivery chain to produce, monitor, and report on meaningful outcomes using metrics that citizens can see and understand (Barber, 2016; Barber et al., 2011a, 2011b; Murekeyisoni, 2017). This specialized delivery unit is to lead the way to cultural change in government and throughout delivery chains such that the institutional buzzwords become "ambition, focus, clarity, urgency and irreversibility" (Barber et al., 2011a; see pp. 31–33).

The second step requires an assessment of past and present performance as well as the drivers of performance. Next, the government must proceed to establish a delivery plan comprising three key components: (1) the reform strategy, (2) the delivery chains, and (3) the targets and trajectories. The metrics must be clear, measurable, and meaningful to citizens and, especially, to politicians to help them "run a government" and usher in change efficiently. Finally, it is necessary to drive delivery by establishing routines for performance monitoring and problem-solving, and by maintaining the momentum. Thus, deliverology seems to address the bounded rationality problem of policymakers by proposing a simple plan: the relentless pursuit of a set of clear goals with the prioritization and tracking of key metrics and the delivery capacity to act quickly when needed.

Barber presented deliverology to the Canadian cabinet as a four-year pledge-fulfillment strategy: "Set the agenda and develop the approach" and "show early wins" (year 1), "establish the delivery discipline across government" and "drive for results" (year 2), "continue drive for results" and "plan delivery of the next parliament's top priorities" (year 3), and "manage through to success and run-up to the election" (year 4) (see Barber, 2016; Canada School of Public Service, 2016a). Journalists, commentators, and some academics tend to perceive deliverology as a pledge-fulfillment tool that helps the government to achieve measurable, concrete outcomes that respect election promises and keeps them accountable to citizens while improving governance through continuous monitoring of implementation (Arellano, 2016a, 2016b, 2016c; Castonguay, 2016; Lahey, 2017; Murekeyisoni, 2017; Wells, 2016a; Wherry, 2016, 2017; Zilio, 2016). Ironically, despite Barber's critique of short-term thinking and government by political communication,

deliverology promotes a management and communication strategy that fits with the electoral logic of the permanent campaign in representative democracy, where governing parties are always preparing for, recovering from, and engaging in elections (Couture & Jacob, 2019).

By defining the election promises as the deliverables and planning for their delivery over the electoral cycle, deliverology implicitly embraces the traditional model of electoral or democratic accountability known as promissory representation, whereby citizens grant a mandate to a political party to govern and hold the government accountable to keeping its election promises (Mansbridge, 2003). Research shows that, deliverology or not, the promise-keeping records at the federal level in Canada and provincially in Quebec are quite good (Pétry, 2014; Pétry & Birch, 2016, 2018; Pétry & Duval, 2015). The governing party seeks to satisfy voter preferences expressed by the election results. Recent work examining the performance of the Couillard government in Quebec (2014–2018) and the Trudeau government (2015–present) shows that while pledge fulfillment is an important aspect of governance between elections, much legislative and policy action is not linked directly to election promises but rather to agenda-setting dynamics for new issues and policy-cycle dynamics for existing policies and programs (Jacob, Birch, Pétry, & Baby-Bouchard, 2018; Birch, Jacob & Baby-Bouchard, 2019). Furthermore, in a pioneering study, Werner (2018) discovered that when given a choice between democratic representation models, fewer than one in ten respondents preferred promissory representation through faithful promise keeping, while half chose the trustee or anticipatory model of governing for the common good (see also Mansbridge, 2003, for a discussion of all representation models). Without a mutually exclusive choice between models (i.e., with separate questions about each model), respondents gave each representation model high importance rankings. Werner concludes that voters "want it all"—governing parties that are promise keepers and trustees of the common good and that show responsiveness to public opinion. In light of these findings regarding democratic accountability, it is necessary to ask what the value-added of deliverology is, as a way to implement election promises deemed to be of inherent worth, when governments are already quite good at promise keeping and when citizens may expect more than promise keeping.

HOW DOES DELIVEROLOGY COMPARE TO EVALUATION AND RESULTS-BASED MANAGEMENT?

On the surface, deliverology seems to include a form of monitoring or concomitant evaluation (as opposed to *ex ante* or *ex post* evaluation), with its emphasis on tracking and targeting key metrics. These metrics are similar to results-based targets in public-sector management. Tweedie (2018), a policy developer and strategist in the government of Canada, argues that deliverology is really a rebranding of results-based management. Deliverology supports recourse to dashboards, but with a small set of specific targets deemed to be meaningful for citizens.

Both evaluation and deliverology share an ambition to ensure social betterment through the development, design, and implementation of effective policies and programs that yield meaningful outcomes. However, deliverology focuses on implementing the governing party's vision of betterment via a delivery unit, while the multiple evaluation units dispersed throughout government agencies question the coherence between policy aims, implementation, and outcomes, their appropriateness, and the soundness of underlying logic models. As [Table 1](#) shows, there are many substantive distinctions between the two worlds. Additional differences concern the vocabulary and scope of modern theory-based evaluations that attend to both the intended and unintended outcomes, using a variety of qualitative and quantitative data sources rather than a shortlist of key metrics.

WHAT ARE THE CRITICISMS OF DELIVEROLOGY AND ITS CONSEQUENCES FOR EVALUATION AND EVIDENCE-INFORMED POLICYMAKING?

Positive perspectives on deliverology start from Barber's claims that it enhances public-sector management, performance, and accountability to better serve citizens, regardless of the conditions at the outset of adopting deliverology. [Barber \(2016\)](#) and [Castonguay \(2016\)](#) suggest that deliverology can alleviate taxpayer dissatisfaction and improve confidence in government. However, most of these claims are prospective opinions that have not been verified empirically through a rigorous independent evaluation process.³ Robert [Lahey \(2017, p. 7\)](#), a Canadian evaluation expert and practitioner turned consultant, who was the founding head of Canada's Centre of Excellence for Evaluation, noted, "On the positive side, deliverology, given its link to political power, generally brings with it authority, resources, flexibility and a striving for provision of timely advice and quick turnaround (i.e. a sense of urgency that can potentially cut through bureaucratic roadblocks to action)."

In one of the rare independent case studies of the concept, [Murekeyisoni \(2017\)](#) finds positive effects in her assessment of deliverology applied to Ontario's education reform in the McGuinty years. She found that deliverology facilitated efficient management and coordination while ensuring a citizen-centered, results-based approach to educational reform. She further noted that it helped address resistance to change in the delivery chain. However, this success reflects adaptations of the deliverology model to engage actors collaboratively in defining goals and targets, rather than a standard top-down delivery pattern, and to grant flexibility to actors to adapt to change.

Many negative criticisms situate deliverology as a variant of results-based management and refer to research findings about the pitfalls of the application of private-sector-inspired reforms with performance targets in the public sector ([Hood & Dixon, 2015a, 2015b](#); [Manning, 2015](#)). For example, deliverology is criticized for accentuating the hierarchical concentration of power within the Prime Minister's Office (PMO), to the detriment of accountability

Table 1. Contrasting deliverology and evaluation

Comparison point	Deliverology	Evaluation
Purpose	Promotion and implementation of delivery units as a method of ensuring policy and reform delivery to improve government functioning to benefit the lives of citizens	Production of valid and reliable knowledge to decision makers about programs and organizations with a view to social and organizational betterment
Target audience for reporting	Prime Minister, Cabinet, and the electorate	Managers and executive decision makers within government
Type of accountability	External accountability	Internal accountability
Focal point(s)	Fulfillment of government's agenda, election promises, and public-sector reforms	Fulfillment of stakeholders' needs for data about relevance, accessibility, comprehensiveness, integration, fulfillment of objectives, effectiveness, impact, cost, efficiency, and sustainability (OECD, 2010; Patton, 1997).
Evidence base—main sources	Practitioner experiences and stories of successful case studies about implementation	Scientific and evaluative knowledge gathered using social research methods to determine how and why what works and does not work for whom in different contexts
Perspective on evidence and policy	Discursive support of the idea of "evidence-inspired" policy	Implicit and explicit commitment to the idea of evidence-informed policy
Approach to data analysis	Tracking of selected metrics for reaching targets for each government priority to assure "relentless implementation"	Analyzing multiple quantitative and qualitative data sources to ensure a complete picture
Approach to identifying indicators	Considering metrics chosen by the executive with the Delivery unit as inherently appropriate	Using logic theories to identify relevant standards or performance indicators and to focus on overall intended outcomes
Time horizon	Immediate and short-term assessments during an election mandate	Multiple assessments of outcomes and impacts over time, with a focus on the medium- and long-term impact
Adherence to standards	Ensuring transparency about delivery of the governing party's pledges	Respecting formal program evaluation standards (Joint Committee on Standards for Educational Evaluation) and ethics principles

Note. Information adapted from Barber et al. (2011a, 2011b), Canadian Evaluation Society (2014)

mechanisms such as institutionalized evaluation in the public service and reporting to Parliament. Top-down command-and-control approaches tend to demotivate public-service workers by failing to recognize their existing expertise and by considering them as obstacles to policy reform rather than partners in delivery. Empirical work analyzing 20 agencies operating with performance contracts in Quebec revealed that better outcomes arose when decision-authority and financial controls were decentralized, not centralized (Aubert & Bourdeau, 2012), which is in keeping with Murekeysoni's (2017) comments on the adaptations of deliverology in Ontario. In contrast to Murekeysoni's and Barber's own claims about the effectiveness of deliverology for Ontario's education reforms, Sattler (2012) found that these education-governance reforms were rooted in a neoliberal accountability and performance reform agenda starting in 1990 that was implemented through a long-term, incremental, and "messy" process, with no evaluation of the ultimate impact on outcomes for students. Furthermore, some UK case studies also suggest a continuum with pre-existing new public management reforms and expose the limits to deliverology due to short-sighted targets, missed delivery problems, the complexity of policy implementation, and the challenges of changing organizations and target population behaviors in durable ways (Smith, Richards, Geddes, & Mathers, 2011; Smithers, 2001).

A second line of criticism concerns the targets themselves, pressures to reach targets, and unintended, perverse, or short-lived effects. When targets lack legitimacy among practitioners, focus on a narrow set of metrics to the exclusion of others, and include mixed incentives attached to employee performance assessment or to program funding, they can undermine results and outcomes, yielding perverse, unintended results by demotivating and devaluing practitioners, creating a climate of fear about the consequences of missing targets, and by placing them in ethical dilemmas (Arnaboldi, Lapsley, & Steccolini, 2015; Ayers, 2013; Jacob, 2009a, 2009b; Jacob, Miller-Pelletier, & Birch, 2017; Murekeysoni, 2017; Poister, 2003). By focusing on chosen metrics, deliverology may sacrifice the systematic analysis of unintended, even perverse consequences that will arise if actors down the delivery chain find incentives to "present selective data," "game the metrics," and "distort or manipulate the findings" (Bevan, 2012; Bevan & Hood, 2006; Fisher & Downes, 2008; Hood, 2006, 2010, 2012; Schacter, 2002a, 2002b, 2006, 2011, 2016; Welsh & Ordóñez, 2014). The harshest and most sarcastic criticism of deliverology comes from John Seddon (2008, 2013), who argues using examples from deliverology under the Blair government that it is an oversimplified form of "Mickey Mouse command and control" that believes in "reform by targets" but misses the desired policy outcomes. He concludes that deliverology results in a dysfunctional, distorted system where demoralized, cynical service providers "manage all the wrong things," by focusing on costs rather than values and on gaming targets rather than achieving higher outcomes, which leads to increased costs to support gaming strategies, with perverse effects on outcomes especially in the health-care and education sectors (see CFA Local 1983 [2010a, 2010b] for videos of Seddon's speeches condemning deliverology). To the extent

that deliverology is "old wine in a new bottle," these problems with results-based approaches are relevant.

Another consequence of results-based approaches is the trend toward using evaluation expertise in Canadian public administration to develop, track, and report on a limited set of metrics, which diminishes the time and resources allocated to conducting all-encompassing evaluations that ensure accountability and learning for outcomes and do not just target achievements and reporting up the internal hierarchy (Savoie, 2015). Politically driven and externally oriented accountability mechanisms embedded in performance regimes seem to miss the mark for measuring performance and for creating positive incentive systems for enhanced service delivery and policy learning, especially when policies entail complex tasks that are difficult to capture in a limited range of indicators (Jakobsen, Baekgaard, Moynihan, & van Loon, 2018). Already, without the addition of deliverology, this narrow approach undermines the role and production of *ex post* evaluations as a tool for policy learning and improving outcomes (Shepherd, 2018). When delivery units are created within departments in parallel with existing audit and evaluations departments, multiple tensions arise between the functions of delivery, auditing, and evaluation. Delivery units, because of their direct links to the PCO and their preoccupation with results indicators, further institutionalize a form of new political governance that may squeeze audit and evaluation functions. For civil servants, ethical tensions can emerge between the public-sector ethos of neutrality and the perceived partisan prerogatives of delivery units, especially when the line between public service in the government and public service under the governing party becomes fuzzy.

Furthermore, deliverology targets are more limited in range and depth than the indicators of program relevance, effectiveness, and outcomes that evaluators would choose or that may be meaningful to citizens. When deliverology targets are met while other more significant indicators are not, this may actually undermine legitimacy, as citizens perceive a gap between the successful attainment of a target and their own experiences as the beneficiaries of a public service. When the broader range of evaluation criteria and indicators is ignored, it is likely to take more time and be more tedious to improve delivery, beneficiary satisfaction, and desired long-term outcomes. Although command-and-control approaches to delivery may yield short-term results, long-term change requires modifications in organizational culture and behaviour. This may stall as the next election nears because the delivery concerns of politicians metamorphose into a focus on the campaign ahead, and those of civil servants anticipate those of the next government.

Although Barber recognizes the growing criticism of the effectiveness of delivery units, given their variable results, his response is as follows:

This critique is important, but is not completely fair, as the idea of a Delivery Unit has become a victim of its own success. When something is fashionable, lots of people rush to do it, and so impact inevitably varies. As a result, many Delivery Units have the label, but not much else. They belong to a large and growing group of "DINOs"—Delivery Units in Name Only. (Delivery Associates, n.d.b)

A third line of criticism tackles the question of the value-added of deliverology from the perspective of the practitioners of public policy, management, and evaluation as well as of scholars and students of public policy and management. Essentially, deliverology just repackages old ideas, especially but not exclusively about performance management. Many of these ideas, which are standard fare in introductory “101” courses on public policy, administration, or project management, are portrayed as innovative ones by advocates of deliverology. This repackaging is well illustrated in two reports by the Centre for Public Impact, which is co-chaired by Michael Barber, namely “The Public Impact Fundamentals Report: Helping Governments Progress from Idea to Impact” (2016) and “A Rubric for Assessing Public Impact: Applying the Public Fundamentals to the Real World” (2017). These reports present a self-proclaimed “new maximization framework” to close the gap between what governments achieve for their citizens and what they could deliver, a “checklist for practitioners” based on the three public fundamentals (legitimacy, policy, and action), and a rubric for determining what sources of information (such as organization publications, initiative blueprint, initiative implementation plan, speeches from politicians, party manifestos, and public opinion research, plus articles, case studies, and reports) are needed to evaluate the fundamentals. The promotion of these simple, repackaged frameworks and checklists as “new” to professionals who have been applying them and more for some time in their departments errs by ignoring their pre-existing knowledge and expertise. This can foster cynicism and demotivation rather than enthusiasm for the government’s reform agenda. This is more likely when civil servants perceive delivery units as restructuring the space for policy conversations and the roles of existing internal accountability units to their disadvantage or perceive new conflicts between goals, short-sighted targets, and outcomes. Furthermore, according to [Jakobsen et al. \(2018\)](#), external accountability regimes (of which deliverology is one) that are not grounded scientifically and empirically tend to “crowd out,” rather than stimulate, the intrinsic motivation of service providers because of mismatched external incentives and performance indicators, weak consideration of the complexity of policy tasks, and implicit disregard of the importance of intrinsically motivated, autonomous professionals in service management and delivery.

Finally, one may ask whether the “old wine in new bottles” remains, nonetheless, of timeless relevancy or is past its expiry date. This is a question that those who have been promoting the deliverology model relentlessly since the Blair years do not seem to entertain. By contrast, this is the quintessential question for evaluators, public administration scholars, and political scientists. Their theoretical reflections, empirical research, and practitioner experiences regarding results-based management, its organizational and whole-of-government impact, and its relationship to evaluation have supported a trend favoring “*new* wine in new bottles.” This trend goes beyond earlier, simplified, top-down approaches to results-based management to recognize the importance of dialogues between those who make the policy decisions and those who assure implementation. These dialogues involve conversations between agents and principals about performance, the

interface between performance management and evaluation, designing performance systems to encourage organizational learning and effectiveness, policy task complexity, the appropriate indicators, employee motivation, incentive mechanisms as an antidote to cynicism, and the optimal conditions for policy as well as organizational learning (e.g., [Jakobsen et al., 2018](#); [Laihonen & Mäntylä, 2017](#); [Moynihan, 2005, 2008a, 2008b, 2010](#)).

WHAT ABOUT THE IMPLEMENTATION OF DELIVEROLOGY IN CANADA?

Deliverology, like any public management reform, is never implemented in a vacuum. For this reason, let us begin by establishing the context in which it came to Canada through a brief review of historical developments in Canadian public policy and administration. From the 1970s onward, Canada began experimenting with results-based approaches. The country undertook yet another initiative⁴ on "re-thinking government" in the mid-1990s and implemented routines for evaluation, results-based management, regulatory impact assessment, citizen engagement and consultation, stakeholder participation, the production and use of public opinion research, evaluation, and scientific evidence in policy. Although some of these routines were under pressure during the Harper years, they did not disappear from the public-sector landscape. Rather, the Canadian state honed and fine-tuned the infrastructure for monitoring and evaluation and the requirements for the utilization of evaluation results that were instituted with the Treasury Board's Results for Canadians in 2000 and reinforced with mandatory performance frameworks at the program and organizational levels ([Lahey, 2017](#); [Lahey & Nielsen, 2013](#); [Shepherd, 2018](#)). Since 2013, long before deliverology and the mandate letter tracker, the Treasury Board already published online the GC InfoBase, which tracks a variety of indicators and government results ([Canada, Treasury Board Secretariat, n.d.](#)). The 2009 *Evaluation Policy* ([Canada, Treasury Board Secretariat, 2009](#)) that was replaced by the *Policy on Results* ([Canada, Treasury Board Secretariat, 2010](#)) already talked about producing timely, neutral evaluation evidence and managing for results.

In many ways, Canadian public servants have been working within results-based management and accountability frameworks since the 1970s, long before the Trudeau government contracted Barber's services ([Curran, 2016](#); [Lahey, 2017](#); [Lahey & Nielsen, 2013](#); [Tweedie, 2018](#)). In fact, [Tweedie \(2018\)](#) asserts,

We are open, accountable, and transparent. We use plain language and consult with our stakeholders. We have strategic plans, operational plans and implementation plans. We use SMART indicators and ensure that relevance and performance (value for money) drive our organisations to deliver results of benefit to Canadians. We do risk assessments, audits, and evaluations according to clear, published criteria. Modern public policy governance? Nailed it.

However, both [Tweedie \(2018\)](#) and [Savoie \(2015\)](#) point out that the Auditor General of Canada remains critical of public services where the implementation of the

results-based approach resulted in agencies focusing on internal accountability chains, thus neglecting the citizens' perspective and the unintended consequences of the policy in their performance reports ([Auditor General of Canada, 2002, 2004, 2005, 2017](#)). [Lahey and Nielsen \(2013\)](#) also note that the Auditor General supports monitoring and evaluation practices and uptake and frequently calls for more substantive evaluation of overall program effectiveness and quality of results. They also recognize the persistent difficulties the Canadian system experiences in "establishing the right balance" (p. 49) between using performance measurement from monitoring and evaluation for accountability versus policy learning. Given this historical backdrop, deliverology, with its focus on delivering the governing party's agenda and on defining success in terms of reaching targets, may well exacerbate these difficulties by reinforcing this pre-existing trend to favour performance measurement use for control rather than for improving operations through policy learning. Let us now turn to the implementation of deliverology in Canada at the federal level.

On December 23, 2015, Prime Minister Trudeau named Matthew Mendelsohn, former advisor and deputy minister in Ontario's McGuinty government, as Deputy Secretary to the Cabinet (Results and Delivery) in the Privy Council Office. At the first Liberal cabinet retreat in St. Andrews, New Brunswick, in January 2016, deliverology officially appeared on the federal political landscape ([Dyer, 2016a](#)). Sir Michael Barber from Delivery Associates, under contract for two years at \$217,000 ([Castonguay, 2016](#)), Dominic Barton, CEO from McKinsey Consulting, working pro bono ([Geddes, 2016](#)), and their ideas became special guests at this meeting and at subsequent federal cabinet retreats in Kananaskis, Alberta (April 2016), Sudbury, Ontario (August 2016), St. John's, Newfoundland and Labrador (September 2017), and London, Ontario (January 2018). Through the Results and Delivery Unit in the Privy Council Office, Barber's ideas became embedded in the Trudeau government, which created a cabinet committee on "Agenda and Results," chaired by the Prime Minister, which would become the "Agenda, Results and Communication" committee by August 2016 ([Wells, 2016b](#)) and subsume the cabinet committee on open government and parliamentary affairs in August 2018. The Canada School of Public Service offered videos featuring Barber ([Canada School of Public Service, 2016a, 2016b](#)), and the [Institute on Governance \(2018\)](#) offered training packages on deliverology. The Trudeau government deployed considerable effort at the apex of power to show a commitment to effect a deliverology culture shift from the top of the Canadian government down the delivery chains.

By July 1, 2016, the government had introduced the *Policy on Results* and the *Directive on Results* to replace the *Policy on Reporting of Federal Institutions and Corporate Interests* to [Treasury Board of Canada Secretariat \(2007\)](#), the *Policy on Evaluation* (2009), and the *Policy on Management, Resources and Results Structures* (2010) ([Canada, Treasury Board Secretariat, 2016a, 2016b](#)). Under the new policy, the evaluation function remains embedded in the Canadian state, with some new requirements. The *Policy on Results* explicitly links evaluation with performance measurement semantically and structurally through the Performance

Measurement and Evaluation Committee. Departments began creating new positions for a Chief Results and Delivery Officer and a Chief Data Officer.

Then, in November 2017, the Trudeau government launched the ultimate deliverology tool: the Mandate Letter Tracker ([Canada, Privy Council Office, 2018a, 2018b](#); [MacCharles, 2017](#)). In January 2018, Matthew Mendelsohn, head of Canada's delivery unit, presented deliverology as a "data-driven, results-focused approach" to public-sector management that included a space for evidence and evaluation, despite the challenges of "obtaining timely data for metrics on pledge fulfillment" (Mendelsohn, as interviewed by [Haws, 2018](#)). He seems to suggest that the value-added of the Mandate Tracker is to provide greater transparency about pledge tracking, a task that previous governments did behind closed doors, and to focus executive-level attention on implementation through "ongoing evaluation" as opposed to conventional *ex post* evaluations. In contrast, political scientists express serious concerns that deliverology encourages deeper centralization at the executive level with special communication channels and additional resources for delivery units (Alex Marland and Anna Esselment, as interviewed by [Ryckewaert, 2016](#)).

From the start, the Canadian media responded to these developments in federal governance with skepticism, calling Barber a "guru" ([Arellano, 2016c](#); [Dyer, 2016b](#); [McGregor, 2016](#)), the Prime Minister a "devotee" ([Wherry, 2016](#)), and deliverology as the latest "fad and fashion" ([French, 2016](#)) or a "punchline" ([Radwanski, 2017](#)), with growing criticism ([Wherry, 2017](#)) that switched to sarcasm when denied an interview with Barber on broken election promises and the use of "deliverology for relentless implementation" ([Bimman, 2018](#)). Responding to an article in *L'Actualité* ([Castonguay, 2016](#)), the Canadian Evaluation Society reminded readers of the 40 years of evaluation practice in Canada and the institutionalization of evaluation in the federal administration starting in 1977 ([Cummings & Gauthier, 2016](#)). At a press conference to release fall reports, Michael Ferguson, the [Auditor General of Canada \(2017\)](#), signalled that there is a delivery gap between the whole-of-government level and its messages about citizen-centric service delivery. When queried about the mandate tracker, Ferguson remarked that the limited scope of the mandate tracker with its focus on election promises missed the fact that "there's more to government than big-talk mandate letters for ministers" (as cited by [Delacourt, 2017](#)). In many ways, the "Mandate Letter Tracker" website seems to be a new tool for political communications and packaging the government's progress (or lack of progress) toward achieving its election commitments as reformulated and adjusted in the mandate letters.⁵

The arrival of deliverology in Canada sparked a friendly debate between [Mark Schacter \(2016\)](#), an Ottawa-based consultant and author of a critical essay entitled "Does 'Deliverology' Matter?," and [Steve Montague \(2016\)](#), an adjunct professor and consultant who offered "A Response to Does 'Deliverology' Matter?: Targets Can Work If Done Properly." Although Montague recognized many of Schacter's criticisms, he expressed reservations about skepticism toward targets. He argued for the "responsible" development and management of targets

over the longer policy cycle in an iterative process that engages stakeholders and decision makers and contributes to policy learning. Thus, Montague implicitly embraced the “new wine” approach based on performance dialogues for learning and explicitly hoped that the Canadian brand of deliverology would engage in “responsible” targeting.

CONCLUSION

This tale of two worlds traces the portrait of deliverology as a business and a model of implementation that repackages and rebrands old ideas linked to results-based management and democratic accountability as pledge fulfillment, while assuming that current governance systems are not delivering and therefore need a deliverology fix. It questions the innovativeness of deliverology generally and highlights some criticisms of this model, especially regarding the possibility that deliverology may usurp evaluative capacity. It finds limited value-added, since recourse to deliverology in Canada ignores both the strong record of pledge fulfillment and the particularities of Canada’s complex multi-level governance context. Although deliverology does bring a focus on implementation to the apex of power, its relevance and effectiveness seem somewhat elusive in a context where the state has already institutionalized evaluation and results-based approaches and where new political governance undermines frontline delivery capacity via the inverted “90/10” split. The effectiveness of deliverology as a “way to run a government” has never been subjected to an independent, whole-of-government evaluation, and, if it were, it would be difficult to disentangle outcomes owing to deliverology from those owing to other factors in a context of institutionalized evaluation and monitoring systems. Originally a “tongue-in-cheek” term coined by Barber’s critics ([The Economist, 2015](#)), deliverology became Barber’s globalized consulting enterprise. Barber’s deliverology has an aura of scientificity because of recourse to target metrics, and an appealing message because of its practical, business-like approach to complex governance challenges associated with public sector reforms.

Deliverology is light on science and heavy on narrative, with flexible evidence standards that favour “best practice stories” and narrower “targeted metrics,” whereas evaluation and social research are science heavy and narrative light, with stricter evidence standards, assessment criteria, and multiple metric points tailored to specific purposes. Each world has its own value orientation. Barber relies on business-inspired thinking about running a government that inherently values private-sector performance management. His model is shaped by the ideological orientations of the self-proclaimed, “small ‘I’ liberal” and “centre-right” think tanks and organizations that Barber refers to in his works. By contrast, evaluators and policy researchers are grounded in scientific thinking, leading them to value systematic recourse to theories, research methods, and empirical findings to address policy questions, to assess implementation outcomes, and to determine the effects of programs and policies, as compared to the “counterfactual”—that is, what would have occurred had the policy or program not been implemented,

or if they had been implemented differently. To them, many of Barber's sources are "grey literature." To Barber, the work of evaluators and policy scientists is too abstract and lacks synthesis. He cares about helping governments to deliver their election promises and reforms, whatever they may be, abstracted from any critical analysis of the reform agenda itself, the need for structural change within the public administration system, or the current pledge-delivery capacity of the system. By contrast, evaluators and political scientists care about the relevancy, appropriateness, and impact of such reforms—that is, their ability to achieve socially desirable outcomes with target populations in meaningful, cost-effective ways, and their ability to enhance public management and service delivery. Where evaluators aspire to facilitate conversations about "good" commitments and outcome congruence with policy goals, "deliverologists" assume party commitments are "good" and seek conversations about implementation and progress toward achieving set targets.

Perhaps Barber can learn from evaluators, public administration scholars, and political scientists about the importance of context, evidence of the perverse effects of governance by targets, evidence of the failures of private-sector-inspired reforms on public-sector capacity and employee motivation, and, most important, the role of more comprehensive evaluations of reforms, programs, and policies in deep policy learning and delivery of better outcomes. If nothing else, perhaps evaluators, public administration scholars, and political scientists should take note of the importance of "storytelling" and "packaging" as a mode of knowledge diffusion for uptake by policymakers. Barber's success in global networking to diffuse his "Deliverology" model shows that storytelling about the effectiveness of this governance model, even without solid, science-based evidence, is more appealing to policymakers who say they support "evidence-informed" policy and management but who struggle with bounded rationality in complex, challenging policy environments.

ACKNOWLEDGEMENTS

We wish to thank Raphaëlle Jacques for her contributions to the reference list. We acknowledge funding from The Fonds de recherche du Québec—Société et culture (FRQSC) for our research team.

NOTES

- 1 [Savoie \(2015\)](#) asserts that "To those inside, particularly where it matters increasingly in central agencies and in the Prime Minister's Office (PMO), government is 90 per cent ideas and 10 per cent implementation. To those outside, government should be 90 per cent implementation and 10 percent ideas.... There is a fault line in government that separates those responsible for generating new policies and managing the blame game from those responsible for implementation. That explains why government was not as good as it once was in delivering programs and services" (pp. 14–15). Like Savoie, [Michael Barber \(2016, p. 36\)](#) advises politicians that "policy is 10 per cent and implementation is 90 per

- cent.” Oddly, neither Savoie nor Barber references the other’s work, nor do they reference similar literature, despite their shared diagnostic of the 90/10 split and their criticism of the short-sightedness of “government by political communications.”
- 2 The exception is a reference to Osbourne and Gaebler’s famous book *Reinventing Government: How the Entrepreneurial Spirit Is Transforming the Public Sector* (1993), which emphasizes market-oriented, results-oriented governments that “steer” rather than “row.”
 - 3 We began our research efforts and analysis of “deliverology” in 2016 at the same time as others (see Richards, 2018; Richards et al., 2017; Schacter, 2016). Our findings corroborate one another’s. We add that much of the positive literature on “deliverology” in cyberspace is linked to Barber or to one of the consulting firms with which he is or was associated.
 - 4 Since the Glassco commission of 1962–63, the Canadian government has experienced multiple iterative processes that have led to the institutionalization of evaluation and shifts toward evidence-based decision making.
 - 5 Governing parties in Canada have used similar tools to track their promise delivery, without calling it “deliverology.” See the Campbell government’s “Five Great Goals for a Golden Decade” in B.C., the Stelmach government’s “Measuring Up: Progress Report on the Business Plan of the Alberta Government,” the Charest government’s short-lived tracking tool or the traffic light system of the Marois government in Quebec, and the federal government’s “Departmental Results Framework and Inventory Program.”

REFERENCES

- Arellano, N. (2016a). Can deliverology help solve education, health problems in remote areas? *Canadian Government Executive*. Retrieved from <http://canadiangovernmentexecutive.ca/can-deliverology-help-solve-education-health-problems/>
- Arellano, N. (2016b). Trudeau’s five deliverology targets. *Canadian Government Executive*. Retrieved from <http://canadiangovernmentexecutive.ca/trudeaus-five-deliverology-targets/>
- Arellano, N. (2016c). Trudeau’s “deliverology” guy. *Canadian Government Executive*. Retrieved from <http://canadiangovernmentexecutive.ca/trudeaus-deliverology-guy/>
- Arnaboldi, M., Lapsley, I., & Steccolini, I. (2015). Performance management in the public sector: The ultimate challenge. *Financial Accountability & Management*, 31(1), 1–22. <https://doi.org/10.1111/faam.12049>
- Aubert, B. A., & Bourdeau, S. (2012). Public sector performance and decentralization of decision rights. *Canadian Public Administration*, 55(4), 575–598. <https://doi.org/10.1111/j.1754-7121.2012.00238.x>
- Aucoin, P. (2012). “New Political Governance” in Westminster systems: Impartial public administration and management performance at risk. *Governance: An International Journal of Policy, Administration and Institutions*, 25(2), 177–199. <https://doi.org/10.1111/j.1468-0491.2012.01569.x>
- Auditor General of Canada. (2002). Chapter 6: A model for rating department performance reports. In *2002 April Report of the Auditor General of Canada*. Retrieved from https://web.archive.org/web/20190618194158/http://www.oag-bvg.gc.ca/internet/English/parl_oag_200204_e_1133.html

- Auditor General of Canada. (2004). Chapter 7—Managing government: A study of the role of the Treasury Board and its Secretariat. In *2004 March Report of the Auditor General of Canada*. Retrieved from http://www.oag-bvg.gc.ca/internet/English/parl_oag_200403_07_e_14899.html#ch7hd4b
- Auditor General of Canada. (2005). Chapter 5: Rating selected department performance reports. In *2005 Report of the Auditor General of Canada*. Retrieved from http://www.oag-bvg.gc.ca/internet/English/parl_oag_200504_05_e_14936.html
- Auditor General of Canada. (2017, November 21). *Auditor General's opening statement—2017 Fall reports press conference*. Retrieved from http://www.oag-bvg.gc.ca/internet/English/osm_20171121_e_42705.html
- Ayers, R. S. (2013). Building goal alignment in federal agencies' performance appraisal programs. *Public Personnel Management*, 42(4), 495–520. <https://doi.org/10.1177/0091026013496077>
- Barber, M. (2008). *An instruction to deliver: Tony Blair, the public services and the challenge of delivery*. London, England: Politicos.
- Barber, M. (2013). *The good news from Pakistan*. Retrieved from <https://web.archive.org/web/20180823130848/http://www.reform.uk/publication/the-good-news-from-pakistan/>
- Barber, M. (2016). *How to run a government so that citizens benefit and taxpayers don't go crazy*. London, England: Penguin Books.
- Barber, M. (2017). *How to deliver improved outcomes for schools*. Retrieved from <https://www.wise-qatar.org/2017-wise-research-improved-outcomes-school-systems/>
- Barber, M., Donnelly, K., & Rizvi, S. (2013). *An avalanche is coming: Higher education and the revolution ahead*. London, England: Institute for Public Policy Research. Retrieved from <http://www.avalancheiscoming.com/>
- Barber, M., Moffit, A., & Kihn, P. (2011a). *Deliverology 101: A field guide for educational leaders*. Thousand Oaks, CA: Sage.
- Barber, M., Moffit, A., & Kihn, P. (2011b). *Deliverology: From idea to implementation*. Washington, DC: McKinsey and Co.
- Bevan, G. (2012). *The challenge of designing "good enough" performance measures & results framework*. London, England: London School of Economics.
- Bevan, G., & Hood, C. (2006). What's measured is what matters: Targets and gaming in the English public health care system. *Public Administration*, 84(3), 517–538. <https://doi.org/10.1111/j.1467-9299.2006.00600.x>
- Birch, L., Jacob, S., & Baby-Bouchard, A. (2019). The Trudeau government's legislative agenda: Election promises and a dual mandate. In L. Birch & F. Pétry (Eds.), *Assessing Justin Trudeau's Liberal government: 353 promises and a mandate for change* (pp. 27–42). Quebec, QC: Presses de l'Université Laval.
- Bimman, A. (2018, January 12). Analysis: Trudeau cabinet's special guests. *Global News*. Retrieved from <https://globalnews.ca/news/3962858/trudeau-cabinet-retreat-guests/>
- Brown, A. (2015, March 10). *Helping governments govern*. Retrieved from <https://www.bcg.com/publications/2015/public-sector-helping-governments-govern.aspx>

- Cairney, P., & Oliver, K. (2017). Evidence-based policymaking is not like evidence-based medicine, so how far should you go to bridge the divide between evidence and policy? *Health System Policy and Research*, 15(1), 35. <https://doi.org/10.1186/s12961-017-0192-x>
- Canada, Privy Council Office. (2018a). *Mandate letter tracker: Delivering results for Canadians*. Retrieved from <https://www.canada.ca/en/privy-council/campaigns/mandate-tracker-results-canadians.html>
- Canada, Privy Council Office. (2018b). *The mandate letter tracker and the results and delivery approach*. Retrieved from <https://www.canada.ca/en/privy-council/services/results-delivery-unit.html>
- Canada School of Public Service. (2016a). *Results and delivery—Lessons from around the world* [Video file]. Retrieved from <https://www.youtube.com/watch?v=jUJIDAy6KbY>
- Canada School of Public Service. (2016b). *Sir Michael Barber: Results and delivery unit: Lessons for Canada*. Matthew Mendelsohn interviews Michael Barber [Video file]. Retrieved from <https://www.youtube.com/watch?v=JogXHgPXhVs>
- Canada, Treasury Board Secretariat. (2007). *Policy on reporting of federal institutions and corporate interests to Treasury Board Secretariat*. Retrieved from <http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=12622>
- Canada, Treasury Board Secretariat. (2009). *Policy on evaluation*. Retrieved from <http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=15024>
- Canada, Treasury Board Secretariat. (2010). *Policy on management, resources and results structures*. Retrieved from <http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=18218>
- Canada, Treasury Board Secretariat. (2013–2014). *Evaluation of the 2009 policy on evaluation*. Retrieved from <https://www.canada.ca/en/treasury-board-secretariat/services/audit-evaluation/centre-excellence-evaluation/evaluation-2009-policy-evaluation.html>
- Canada, Treasury Board Secretariat. (2016a). *Policy on results*. Retrieved from <http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=31300>
- Canada, Treasury Board Secretariat. (2016b). *Directive on results*. Retrieved from <http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=31306>
- Canada, Treasury Board Secretariat. (n.d.). *Infographic for government*. Retrieved from <https://www.tbs-sct.gc.ca/ems-sgd/edb-bdd/index-eng.html#orgs/gov/gov/infograph/results>
- Canadian Evaluation Society. (2014). *What is evaluation?* Retrieved from <https://evaluationcanada.ca/what-is-evaluation>
- Castonguay, A. (2016, November 28). Voici l'arme secrète de Trudeau. *L'actualité*. Retrieved from <http://lactualite.com/politique/2016/11/28/voici-larme-secrete-de-trudeau/>
- Cantelli, F., Jacob, S., Genard, J.-L., & de Visscher, C. (Eds.). (2006). *Les constructions de l'action publique*. Paris, France: L'Harmattan, Coll. Logiques politiques.
- Centre for Public Impact. (2016). *The public impact fundamentals report: Helping governments progress from idea to impact*. BCG Foundation. Retrieved from <https://publicimpact.blob.core.windows.net/production/2016/10/CPI-FUNDAMENTALS-REPORT.pdf>
- Centre for Public Impact. (2017). *A rubric for assessing public impact: Applying the public impact fundamentals to the real world*. BCG Foundation. Retrieved from <https://>

- publicimpact.blob.core.windows.net/production/2017/03/5388-CPI-Rubric-For-Assessing-Public-Impact-v3.pdf
- Centre for Public Impact. (n.d.). *What is public impact?* BCG Foundation. Retrieved from <https://www.centreforpublicimpact.org/what-is-public-impact/>
- CFA Local 1983. (2010a, April 20). *Why Deliverology made things worse in the UK. Part 2.* Speech by John Seddon at California Faculty Association Assembly [Video file]. Retrieved from <https://www.youtube.com/watch?v=6n9TH5ktKFE>
- CFA Local 1983. (2010b, April 20). *Why Deliverology made things worse in the UK. Part 3.* Speech by John Seddon at California Faculty Association Assembly [Video file]. Retrieved from <https://www.youtube.com/watch?v=2sIFvpRilSc>
- Chouinard, J. A., & Milley, P. (2015). From new public management to new political governance: Implications for evaluation. *Canadian Journal of Program Evaluation*, 30(1), 1–22. <https://doi.org/10.3138/cjpe.30.1.1>
- Couture, J., & Jacob, S. (Eds.). (2019). *Démocratie et politiques publiques*. Quebec, QC: Presses de l'Université Laval.
- Cummings, H., & Gauthier, B. (2016). *La SCÉ éduque l'Actualité à l'évaluation de programme*. Retrieved from <https://evaluationcanada.ca/fr/nouvelles/7306>
- Curran, R. (2016, April 27). Will deliverology work for the federal government? *Policy Options/Options politiques*. Retrieved from <http://policyoptions.irpp.org/magazines/april-2016/is-deliverology-right-for-canada/>
- Delacourt, S. (2017, November 21) *Stuck on hold: The AG's report and a government that can't get things done*. Retrieved from <https://ipolitics.ca/2017/11/21/stuck-hold-ags-report-government-cant-get-things-done>
- Delivery Associates. (n.d.a). *Who we are*. Retrieved from <https://www.deliveryassociates.com/who-we-are>
- Delivery Associates. (n.d.b). *Success delivered: How delivery units make a difference for governments and the citizens they serve*. Retrieved from https://uploads-ssl.webflow.com/59ca37d5fcfbf3000197aab3/5aa1526b45d91c0001b762b4_Inc%20London%20-%20DINO%20Brochure%20Amends%20-%20WEB.pdf
- Dobell, R., & Zussman, D. (2018). Sunshine, scrutiny, and spending review in Canada, Trudeau to Trudeau: From program evaluation and policy to commitment and results. *Canadian Journal of Program Evaluation*, 32(3), 371–393. <https://doi.org/10.3138/cjpe.43184>
- Dyer, E. (2016a, January 19). Justin Trudeau's cabinet retreat stays on track until the end, almost. *CBC News*. Retrieved from <http://www.cbc.ca/news/politics/justin-trudeau-cabinet-retreat-new-brunswick-1.3411213> <http://www.cbc.ca/news/politics/justin-trudeau-cabinet-retreat-new-brunswick-1.3411213>
- Dyer, E. (2016b, April 26). "Deliverology" guru schools Trudeau government for 2nd time at cabinet retreat. *CBC News*. Retrieved from <http://www.cbc.ca/news/politics/deliverology-liberal-cabinet-retreat-1.3553024>
- The Economist. (2015, March 26). Improving government: Delivery man. *The Economist*. Retrieved from <https://www.economist.com/news/books-and-arts/21647263-tony-blairs-deliverology-expert-explains-how-reform-public-services-delivery-man>

- Fisher, C., & Downes, B. (2008). Performance measurement and metric manipulation in the public sector. *Journal of Business Ethics: A European Review*, 17(3), 245–258. <https://doi.org/10.1111/j.1467-8608.2008.00534.x>
- French, R. D. (2016, July 20). Trudeau's challenge: Turning promises into policy. *The Globe and Mail* (Ontario edition), p. A13. Retrieved from <https://www.theglobeandmail.com/opinion/trudeaus-challenge-turning-promises-into-policy/article30829997/>
- Geddes, J. (2016, October 31). Why McKinsey & Co. is helping Ottawa out, pro bono. *Maclean's*. Retrieved from <http://www.macleans.ca/politics/ottawa/why-mckinsey-helping-ottawa-pro-bono/>
- Global Agenda Council on Education. (2016). *Unleashing greatness: Nine plays to spark innovation in education* [White paper]. Retrieved from http://www3.weforum.org/docs/WEF_WP_GAC_Education_Unleashing_Greatness.pdf
- Haws, E. (2018, January 24). Oft critiqued Liberal mandate tracker to become “more useful,” says government “deliverology” chief. *The Hill Times*. Retrieved from <https://www.hilltimes.com/2018/01/24/liberals-mandate-letter-tracker-great-conversation-getting-refresh-momentarily-mendelsohn/131770>
- Hood, C. (2006). Gaming in targetworld: The targets approach to managing British public services. *Public Administration Review*, 66(4), 515–521. <https://doi.org/10.1111/j.1540-6210.2006.00612.x>
- Hood, C. (2010). *The blame game: Spin, bureaucracy, and self-preservation in government*. Princeton, NJ: Princeton University Press.
- Hood, C. (2012). Public management by numbers as a performance-enhancing drug: Two hypotheses. *Public Administration Review*, 72(1), S85–S92. <https://doi.org/10.1111/j.1540-6210.2012.02634.x>
- Hood, C., & Dixon, R. (2015a). What we have to show for 30 years of new public management: Higher costs, more complaints. *Governance*, 28(3), 265–267. <https://doi.org/10.1111/gove.12150>
- Hood, C., & Dixon, R. (2015b). *A government that worked better and cost less? Evaluating three decades of reform and change in UK Central Government*. Oxford: Oxford University Press.
- Institute on Governance. (2018). *Results and performance bundle*. Retrieved from <https://web.archive.org/web/20190618163103/https://iog.ca/leadership-learning/course-bundles/results-and-performance-bundle>
- Jacob, S. (2006). Trente ans d'évaluation de programme au Canada : l'institutionnalisation interne en quête de qualité. *Revue française d'administration publique*, 119(3), 515–532. <https://doi.org/10.3917/rfap.119.0515>
- Jacob, S. (2009a). Opération chloroforme ou la réinvention de l'État rationnel : l'évaluation et les données probantes. *Criminologie*, 42(1), 201–223. <https://doi.org/10.7202/029813ar>
- Jacob, S. (2009b). Les systèmes de gestion de la performance et les mutations de la gestion publique contemporaine : enjeux et défis de la logique du classement. *Éthique publique : Revue internationale d'éthique sociétale et gouvernementale*, 11(1), 140–147. <https://doi.org/10.4000/ethiquepublique.1346>

- Jacob, S., Birch, L., Pétry, F., & Baby-Bouchard, A. (2018). Le gouvernement Couillard en tant que "mandataire indépendant." In F. Pétry & L. Birch (Eds.), *Le gouvernement Couillard: 158 promesses et un mandat contrasté* (pp. 77–87). Quebec, QC: Presses de l'Université Laval.
- Jacob, S., Miller-Pelletier, A., & Birch, L. (2017). Les enjeux éthiques de la gestion par les résultats : les enseignements de la loi américaine No Child Left Behind. In B. Mazoux (Ed.), *Gestion par résultats : Concepts et pratiques de gestion de la performance des organisations de l'État* (pp. 173–192). Quebec, QC: Les Presses de l'Université du Québec.
- Jakobsen, M. L., Baekgaard, M., Moynihan, D. P., & van Loon, N. (2018). Making sense of performance regimes: Rebalancing external accountability and internal learning. *Perspectives on Public Management and Governance*, 1(2), 127–141. <https://doi.org/10.1093/ppmgov/gvx001>
- Lahey, R. (2017). Connecting the dots between M, E, RBM ... and deliverology. *Canadian Government Executive*, 23(1), 6–8. Retrieved from http://canadiangovernmentexecutive.ca/dig/23_01/files/assets/basic-html/page-8.html
- Lahey, R., & Nielsen, S. B. (2013). Rethinking the relationship among monitoring, evaluation, and results-based management: Observations from Canada. In S. B. Nielsen & D. E. K. Hunter (Eds.), *Performance Management and Evaluation. New Directions for Evaluation*, 137, 45–56. <https://doi.org/10.1002/ev.20045>
- Laihonen, H., & Mäntylä, S. (2017). Principles of performance dialogue in public administration. *International Journal of Public Sector Management*, 30(5), 414–428. <https://doi.org/10.1108/ijpsm-09-2016-0149>
- MacCharles, T. (2017, November 14). Trudeau government's mandate tracker lets you see if federal ministers are keeping commitments. *The Star*. Retrieved from <https://www.thestar.com/news/canada/2017/11/14/trudeau-governments-mandate-tracker-lets-you-see-if-federal-ministers-are-keeping-commitments.html>
- Manning, N. (2015). Supporting "small" improvements in government functioning: A rough guide for development professionals. In *A governance practitioner's notebook: Alternative ideas and approaches* (pp. 151–189). Paris, France: OECD. Retrieved from <https://www.oecd.org/dac/accountable-effective-institutions/Governance%20Notebook%202.5%20Manning%20Small.pdf>
- Manning, N., & Watkins, J. (2013). *Targeting results, diagnosing the means: Innovative approaches for improving public sector delivery*. Recently Asked Questions Series. World Bank. Retrieved from <http://documents.worldbank.org/curated/en/136171467986259116/pdf/98815-BRI-VC-ADD-SERIES-PUBLIC-Box393182B.pdf>
- Mansbridge, J. (2003). Rethinking representation. *American Political Science Review*, 97(4), 515–528. <https://doi.org/10.1017/s0003055403000856>
- McGregor, G. (2016, May 24). Liberal government paid nearly \$18,000 for "deliverology" guru. *CTV News*. Retrieved from <https://www.ctvnews.ca/politics/glen-mcgregor-under-embargo/liberal-government-paid-nearly-18-000-for-deliverology-guru-1.2914536>
- Montague, S. (2016). *A response to does "deliverology" matter?: Targets can work if done properly*. Retrieved from <https://www.pmn.net/wp-content/uploads/Deliverology-Response-June-2016.pdf>

- Moynihan, D. P. (2005). Goal-based learning and the future of performance management. *Public Administration Review*, 65(2), 203–216. <https://doi.org/10.1111/j.1540-6210.2005.00445.x>
- Moynihan, D. P. (2008a). *The dynamics of performance management: Constructing information and reform*. Washington, DC: Georgetown University Press.
- Moynihan, D. P. (2008b). Advocacy and learning: An interactive-dialogue approach to performance information use. In W. Van Dooren & S. Van de Wall (Eds.), *Performance information in the public sector: How it gets used* (pp. 24–41). London, England: Palgrave Macmillan.
- Moynihan, D. P. (2010). A workforce of cynics? The effects of contemporary reforms on public service motivation. *International Public Management Journal*, 13(1), 24–34. <https://doi.org/10.1080/10967490903547167>
- Murekeyisoni, A. (2017). *La résultatologie (Deliverology) comme instrument d'action publique: un cas ontarien* [Master's thesis]. École d'études politiques, Université d'Ottawa, Ottawa, Ontario. Retrieved from <https://ruor.uottawa.ca/handle/10393/36086>
- Nilsen, P., Ståhl, C., Roback, K., & Cairney, P. (2013). Never the twain shall meet? A comparison of implementation science and policy implementation research. *Implementation Science*, 8(1), 63. <https://doi.org/10.1186/1748-5908-8-63>
- OECD. (2010). *Quality standards for development evaluation*. Paris, France: OECD.
- Osborne, D., & Gaebler, T. (1993). *Reinventing government: How the entrepreneurial spirit is transforming the public sector*. New York, NY: Addison-Wesley.
- Patton, M. Q. (1997). *Utilization-focused evaluation: The new century text* (3rd ed.). Thousand Oaks, CA: SAGE.
- Pétry, F. (2014). A tale of two perspectives: Election promises and government actions in Canada. In E. Gidengil and H. Bastedo (Eds.), *Canadian democracy from the bottom up* (pp. 231–252). Vancouver, BC: UBC Press.
- Pétry, F., & Birch, L. (2016). Justin Trudeau's promises, one year later. *Policy Options/Options politiques*. Retrieved from <http://policyoptions.irpp.org/magazines/october-2016/justin-trudeaus-promises-one-year-later/>
- Pétry, F., & Birch, L. (Eds.). (2018). *Le gouvernement Couillard: 158 promesses et un mandat contrasté*. Quebec, QC: Presses de l'Université Laval.
- Pétry, F., & Duval, D. (2015). To what extent do political parties follow their campaign promises? *Journal of Parliamentary and Political Law*. Édition spéciale, 341–364.
- Poister, T. (2003). *Measuring performance in public and nonprofit organizations*. San Francisco, CA: Jossey-Bass, John Wiley & Sons.
- Puttick, R., & Ludlow, J. (2013). *Standards of evidence: An approach that balances the need for evidence with innovation*. NESTA (National Endowment for Science, Technology and the Arts). Retrieved from https://media.nesta.org.uk/documents/standards_of_evidence.pdf
- Radwanski, A. (2017, November 17). Trudeau's "deliverology" on the verge of becoming a punchline. *The Globe and Mail*. Retrieved from <https://www.theglobeandmail.com/opinion/trudeaus-deliverology-on-the-verge-of-becoming-a-punchline/article37023338/>

- Richards, G. (2018). *Does deliverology deliver?* Unpublished Paper. Teffler School of Management, University of Ottawa.
- Richards, G., Gallo, C., & Kronick, M. (2017, January 30). *Does "deliverology" deliver?* Retrieved from <https://canadiangovernmentexecutive.ca/does-deliverology-deliver/>
- Ryckewaert, L. (2016, June 6). Privy Council's new delivery unit increase capacity for centralized control, say experts. *The Hill Times*. Retrieved from <https://www.hilltimes.com/2016/06/06/new-pco-delivery-unit-increases-capacity-for-centralized-control-say-experts/67186>
- Sattler, P. (2012). Educational governance reforms in Ontario: Neoliberalism in context. *Canadian Journal of Educational Administration and Policy*, 128, 1–28.
- Savoie, D. (2015). *What is government good at? A Canadian answer*. Montreal, QC: McGill-Queen's University Press.
- Schacter, M. (2002a). *Not a "tool kit"; Practitioner's Guide to measuring the performance of public programs*. Retrieved from <https://www.files.ethz.ch/isn/122186/Guide.pdf>
- Schacter, M. (2002b). *What will be, will be: The challenge of applying results-based thinking to policy*. Retrieved from <https://iog.ca/research-publications/publications/what-will-be-will-be-the-challenge-of-applying-results-based-thinking-to-policy/>
- Schacter, M. (2006). *Trashing results-based management or throwing out the baby with the bath water*. Retrieved from https://docs.wixstatic.com/ugd/dadb01_1a72e3b10b1e968a8943337920ad8e64.pdf
- Schacter, M. (2011). *Tell me what I need to know: A practical guide to program evaluation for public servants*. Retrieved from https://docs.wixstatic.com/ugd/dadb01_562bd8d94a9b0ca329730d21389d9b3e.pdf
- Schacter, M. (2016). *Does "deliverology" matter?* Retrieved from https://docs.wixstatic.com/ugd/dadb01_67929f50ebaa46efa6e47df3526904a5.pdf
- Seddon, J. (2008). *Systems thinking in the public sector: The failure of the reform regime ... and a manifesto for a better way*. Axminster, England: Triarchy Press.
- Seddon, J. (2013). *Freedom from command and control: Rethinking management for a lean service*. New York, NY: Productivity Press.
- Shepherd, R. P. (2018). Expenditure reviews and the federal experience: Program evaluation and its contribution to assurance provision. *Canadian Journal of Program Evaluation*, 32(3), 347–370. <https://doi.org/10.3138/cjpe.43180>
- Smith, M. J., Richards, D., Geddes, A., & Mathers, H. (2011). Analysing policy delivery in the United Kingdom: The case of street crime and anti-social behaviour. *Public Administration*, 89(3), 975–1000. <https://doi.org/10.1111/j.1467-9299.2010.01892.x>
- Smithers, A. (2001). Education policy. In Anthony Seldon (Ed.), *The Blair effect* (pp 405–426). London, England: Little Brown. Retrieved from <https://www.buckingham.ac.uk/wp-content/uploads/2010/10/blaireduc.pdf>
- Tweedie, J. (2018, June 5). Does "deliverology" deliver the goods? *Canadian Government Executive*. Retrieved from <https://canadiangovernmentexecutive.ca/does-deliverology-deliver-the-goods/>
- Wells, P. (2016a, February 18). Meet Sir Michael Barber, the political delivery man. *Maclean's*. Retrieved from <http://www.macleans.ca/politics/ottawa/meet-sir-michael-barber-the-political-delivery>

- Wells, P. (2016b, August 29). Cabinet committee changes give clues about happenings inside the Trudeau government. *The Hill Times*. Retrieved from <https://www.hilltimes.com/2016/08/29/cabinet-committee-changes-give-clues-about-happenings-inside-trudeau-government/78576>
- Welsh, D. T., & Ordóñez, L. D. (2014). The dark side of consecutive high performance goals: Linking goal setting, depletion, and unethical behavior. *Organizational Behavior and Human Decision Processes*, 123(2), 79–89. <https://doi.org/10.1016/j.obhdp.2013.07.006>
- Werner, A. (2018). Voters' preferences for party representation: Promise-keeping, responsiveness to public opinion or enacting the common good. *International Political Science Review*, 1–16. <https://doi.org/10.1177/0192512118787430>
- Wherry, A. (2016, August 27). How Justin Trudeau plans to deliver on “deliverology.” *CBC News*. Retrieved from <http://www.cbc.ca/news/politics/wherry-trudeau-deliverology-1.3735890>
- Wherry, A. (2017, November 14). Trudeau's Liberals check their to-do list: 67 promises down, more than 200 still to go. *CBC News*. Retrieved from <http://www.cbc.ca/news/politics/liberals-mandate-promises-deliverology-analysis-wherry-1.4400036>
- Zilio, M. (2016, September 30, updated May 18, 2018). British “deliverology” expertise costing Liberals \$200,000. *The Globe and Mail*. Retrieved from <https://www.theglobeandmail.com/news/politics/liberals-spend-200000-for-advice-on-delivering-campaign-pledges/article32187629/>

AUTHOR INFORMATION

Lisa Birch is executive director of the Center for Public Policy Analysis, associate professor in the Department of Political Science at Université Laval, and professor at Champlain St. Lawrence College (Quebec City). Her expertise focuses on public management and public policy, textual analysis, and the fulfillment of campaign promises, including the methodology of the Polimetre, of which she is a co-founder.

Steve Jacob is a full professor in the Department of Political Science at Université Laval and director of the Center for Public Policy Analysis and PerfEval, a research laboratory on public policy performance and evaluation. Trained as a political analyst and historian, he conducts research dealing with the mechanisms of performance management and evaluation: professionalization, institutionalization and capacity building in Europe and Canada, and ethics in evaluation and participatory approaches. The results of his research have been published in numerous journals.

Introduction to Articles Prepared by CES Calgary 2018 Keynote Panel Members on *Reconciliation and Culturally Responsive Evaluation— Rhetoric or Reality?*

Debbie DeLancey
2018 CES Conference Co-Chair

When selecting a theme for the 2018 Canadian Evaluation Society Annual Conference, the co-chairs wanted to build on the CES Board's 2016 motion in response to the Truth and Reconciliation Commission's Calls to Action, committing to including reconciliation in its values, principles, and practices, with a specific commitment to strengthen the promotion of, and support for, culturally responsive evaluation. We wanted a theme that would challenge evaluators to explore what it means to truly work as collaborative partners and allies, not only with Indigenous governments and communities but also with other populations and communities that have traditionally been disempowered.

As conference organizers, we wanted to set the stage with an opening keynote session that would breathe life into the conference theme, would go beyond comfortable platitudes about reconciliation and co-creation, and would pose tough questions about what reconciliation looks like in action.

We recognized that while making a commitment to reconciliation is an important first step, evaluation in Canada still fails to systematically respect Indigenous values and worldviews; it privileges and takes for granted the legitimacy of Western knowledge systems and theories while discounting Indigenous ways of knowing and being. The focus of reconciliation in Canada is on relations between Indigenous and non-Indigenous peoples, but we recognized that these same dynamics of privilege influence our profession's approach to evaluations that deal with other cultures or population groups as well.

For this opening keynote panel, we sought to bring together Canadian and international perspectives on Indigenous evaluation, with speakers who would challenge evaluators to acknowledge traditional power relationships, to consider whether the concept of culturally responsive evaluation is sufficient to advance reconciliation, and to explore what is required for evaluation to effectively advocate for social justice and to support Indigenous sovereignty. Our presenters were asked to address whether the concept of culturally responsive evaluation goes far enough toward advancing reconciliation, to describe successful examples of co-creation, and to leave participants with questions to ponder during the conference. But most importantly, the panel was asked to challenge conference participants to reflect on their own privilege and their

own evaluation practice and consider what changes might be required in the spirit of reconciliation.

Discussions of privilege can be uncomfortable, and even threatening, for some people. As conference organizers, we asked participants to use this opportunity to reflect on their discomfort, consider its source, and be open to the challenges put forward by our keynote presenters.

The opening keynote panel for the CES 2018 Conference was designed to set the stage for three days of exploration and reflection. Larry Bremner, Nicole Bowman-Farrell, Kate McKegg, and Nan Wehipeihana did a brilliant job, challenging conference participants to explore new perspectives, and inspiring all those present. These articles will allow a broader audience to benefit from their insights.

Creating New Stories: The Role of Evaluation in Truth and Reconciliation

Larry Bremner

Proactive Information Services Inc.

Abstract: *This paper describes the origins of the Truth and Reconciliation Commission of Canada, with the focus on how evaluators and their professional associations can contribute to truth and reconciliation. At the professional association level, the actions that the Canadian Evaluation Society has taken in committing itself to incorporating truth and reconciliation into its values, principles, and practices are highlighted. At the individual level, evaluators are challenged to reflect on their practice. As storytellers, evaluators have been complicit in telling stories that, while highlighting the damaging legacy of residential schools, have had little influence on changing the status quo for Indigenous peoples and communities. The need to reconsider who should be telling the stories and what stories should be told are critical issues upon which evaluators must reflect. The way forward also needs to include a move toward a more holistic view, incorporating the interaction between human and natural systems, thus better reflecting an Indigenous, rather than a Western, worldview. The imperative for evaluators, both in Canada and globally, to see Indigenous peoples “as creators of their own destinies and experts in their own realities” is essential if evaluation is to become “a source of enrichment . . . and not a source of depletion or denigration.”*

Keywords: *competency domains, Indigenous approaches, reconciliation, relationality, residential schools, storytelling, Tribal Critical Theory, truth*

Résumé : *Cet article décrit les origines de la Commission de vérité et réconciliation du Canada, en mettant l'accent sur la façon dont les évaluateurs et évaluatrices et leurs associations professionnelles peuvent contribuer à la vérité et à la réconciliation. Il est notamment question des actions que la Société canadienne d'évaluation a posées en s'engageant à inclure la vérité et la réconciliation dans ses valeurs, ses principes et ses pratiques. Au niveau individuel, les évaluateurs et les évaluatrices sont invité.e.s à réfléchir à leur pratique. Comme conteurs et conteuses, les évaluateurs et évaluatrices ont été complices en racontant des histoires qui, tout en soulignant l'héritage néfaste des pensionnats, ont eu peu d'influence sur le changement du statu quo des peuples et communautés autochtones. Il est d'une importance critique que les évaluateurs et évaluatrices réfléchissent à qui devrait raconter les histoires et quelles histoires devraient être racontées. Pour aller de l'avant, il faudra adopter une approche holistique, qui incorpore l'interaction entre l'humain et les systèmes*

Corresponding author: Larry Bremner, P.O. Box 439, Sandy Hook, MB, R0C 2W0; larry@proactive.mb.ca

naturels, reflétant ainsi davantage un point de vue autochtone plutôt qu'occidental. Il sera impératif pour les évaluateurs et évaluatrices, autant au Canada qu'ailleurs dans le monde, de voir les peuples autochtones comme des « créateurs de leur propre destinée et experts de leur propre réalité » si l'évaluation espère devenir « une source d'enrichissement... et non une source d'appauvrissement et de dénigrement. »

Mots clé : domaines de compétence, approches autochtones, réconciliation, relationnel, pensionnats, récits, théorie critique tribale, vérité

[Larry Bremner is a Métis man whose great-grandmother Rose Boucher was born in 1867 in St. Francis Xavier, Manitoba. She moved with her parents by ox team to St. Louis, Saskatchewan, in 1882. In 1883, she married Moise Bremner. On November 19, 1883, Moise, his father, William, and 28 other Métis signed a petition protesting the 1883 Order in Council transferring the Métis lands at St. Louis to the Prince Albert Colonization Company; the petition was ignored by the Canadian government. Moise was a member of Captain Baptiste Boucher's company, one of the 19 *dizaines* (groups of 10 people) led by Gabriel Dumont during the 1885 Métis Resistance. After the resistance at Batoche, the family moved to the United States and returned to what is now Saskatchewan after the Canadian government granted amnesty. They homesteaded in Domremy, Saskatchewan, in 1905.]

INDIAN RESIDENTIAL SCHOOLS IN CANADA

In Canada, residential schools operated until the late 1990s under the auspices of the Government of Canada in partnership with a variety of churches, which included the Roman Catholic, Anglican, United, Methodist, and Presbyterian churches. The first residential school opened in Alderville, Ontario, in 1849, with the last residential school operated by the Canadian government, the Gordon Indian Residential School in Saskatchewan, closing in 1996. It is estimated that over 150,000 Indian, Métis, and Inuit students attended these schools. Many of the children lived in poor conditions and were often sexually and physically abused. Residential schools included “industrial schools, boarding schools, homes for students, hostels, billets, residential schools, residential schools with a majority of day students, or a combination of any of the above. At the request of Survivors, this definition has evolved to include convents, day schools, missions, sanatoriums and settlement camps” ([Legacy of Hope Foundation, 2014](#), p. 3).

Residential schools were intended to assimilate Indigenous peoples into the mainstream dominant culture. Children were forcibly removed from their families, isolated from their homes and communities, and not allowed to speak their language or engage in their culture or traditions, “based on the assumption Aboriginal cultures and spiritual beliefs were inferior and unequal. Indeed, some sought, as it was infamously said, ‘to kill the Indian in the child’” ([Indigenous-Foundations.arts.bc.ca, 2009](#)). The establishment and operation of residential schools in Canada was a vital element in Canada's approach to Indigenous peoples, an approach that has been described as “cultural genocide”:

Cultural genocide is the destruction of those structures and practices that allow the group to continue as a group. States that engage in cultural genocide set out to destroy the political and social institutions of the targeted group. Land is seized, and populations are forcibly transferred and their movement is restricted. Languages are banned. Spiritual leaders are persecuted, spiritual practices are forbidden, and objects of spiritual value are confiscated and destroyed. And, most significantly to the issue at hand, families are disrupted to prevent the transmission of cultural values and identity from one generation to the next.

In its dealing with Aboriginal people, Canada did all these things. (TRC, 2015)

In May 2006, the Indian Residential School Settlement Agreement was approved by all federal political parties. This agreement was seen as an important step in helping to heal the harm caused by the residential school legacy.

The Settlement Agreement represents the consensus reached between legal counsel for former students, legal counsel for the Churches, the Assembly of First Nations, other Aboriginal organizations and the Government of Canada. The implementation of this historic agreement brings a fair and lasting resolution to the legacy of Indian Residential Schools. (Indian and Northern Affairs Canada, n.d.)

THE TRUTH AND RECONCILIATION COMMISSION OF CANADA

Article Seven of the Indian Residential School Agreement called for the establishment of a Truth and Reconciliation Commission. The Truth and Reconciliation Commission of Canada (TRC) was officially launched in 2008. The original Commissioners resigned and the three new Commissioners were appointed: Justice Murray Sinclair, an Ojibwa judge from the court of the Queen's Bench, Manitoba; lawyer Chief Wilton Littlechild from Maskwacis (Hobbema), Alberta; and Marie Wilson, a Canadian Broadcasting Corporation broadcaster from Yellowknife, Northwest Territories. They were officially recognized in a ceremony that took place in July 2009. As noted in the Truth and Reconciliation Commission of Canada Interim Report (2012), the mandate was to "reveal the complete story of Canada's residential school system, and lead the way to respect through reconciliation ... for the child taken, for the parent left behind."

Between 2010 and 2014, the Commission held seven national events. Over 9,000 residential school survivors registered to attend these events, and it is presumed that more attended than registered. Furthermore, it is estimated that over 155,000 individuals attended the national events. In addition, regional events and "town halls" were organized, as well as 238 local hearings in 77 communities across Canada:

Until the Commission was established, the voices of those who were most directly affected by the residential school experience, particularly the former students, had largely been missing from the historical record. The Commission made a commitment to offer everyone involved with the residential school system the opportunity to speak about

their experience. The Commission received over 6,750 statements from Survivors of residential schools, members of their families, and other individuals who wished to share their knowledge of the residential school system and its legacy. (TRC, 2015)

The impact of residential schools has been transferred from grandparents to parents to children, resulting in intergenerational trauma. There are approximately 80,000 residential school survivors still living, so, for many, their stories still need to be heard.

In December 2015, the Truth and Reconciliation Commission of Canada released its Final Report, which included 94 Calls to Action. Many of the Calls to Action, which focus on child welfare, education, health, and justice, call for regular monitoring and evaluation. In addition, the Parliament of Canada was urged to enact a National Council for Reconciliation, with a mandate that was to include the following:

- i. Monitor, evaluate, and report annually to Parliament and the people of Canada on the Government of Canada's post-apology progress on reconciliation to ensure that government accountability for reconciling the relationship between Aboriginal peoples and the Crown is maintained in the coming years.
- ii. Monitor, evaluate, and report to Parliament and the people of Canada on reconciliation progress across all levels and sectors of Canadian society, including the implementation of the Truth and Reconciliation Commission of Canada's Calls to Action. (TRC, 2015)

THE CANADIAN EVALUATION SOCIETY'S RESPONSE TO THE TRC

The Calls to Action hold major implications for all Canadians, including evaluators, and the roles we can play in the process of truth and reconciliation. The Canadian Evaluation Society (CES), the organization representing evaluators in Canada, has taken a few first steps in an attempt to address reconciliation. The CES is committed to incorporating reconciliation in its values, principles, and practices. In May 2016, the Diversity Working Group put forward the following resolutions, which were adopted by the CES National Board at its May 4, 2016, meeting:

- Resolved, That the notion of reconciliation be included in the existing CES value of inclusiveness and that the CES make a public statement to that effect,
- Resolved, That, as part of its next competency scheme review, the CES include reconciliation explicitly,
- Resolved, That the CES strengthen its promotion of and support for culturally responsive evaluation, and

Resolved, That the Diversity Working Group supports the CES in implementing consideration for reconciliation in its activities. (CES, 2016a)

The Diversity Working Group was established in 2008 to advise the CES Board on issues relating to diversity and inclusion within CES and to promote inclusive evaluation practice within the evaluation community.

As the initial step in addressing the first resolution, in June 2016 the CES National President sent a letter to the Minister of Indigenous and Northern Affairs Canada, outlining the resolutions, and stated that “CES will be actively pursuing actions to include these values in its strategies, plans, and activities” (CES, 2016b). The Prime Minister was copied. Other public statements regarding the resolutions and CES’s intent to address reconciliation have been made in presentations at CES and AEA conferences. Furthermore, as evidenced in the recent national conferences, CES has attempted to increase the profile and understanding of Indigenous perspectives as part of its conferences.

Evidence of CES addressing the second resolution can be found in the new competencies that resulted from the CES Credential Competency Review, which included Indigenous input and feedback. There are five competency domains, with multiple competencies within each domain: Reflective Practice, Technical Practice, Situational Practice, Management Practice, and Interpersonal Practice. As part of the competency review process, a number of the original competencies were revised and some new competencies were added. The following are examples of a new and a revised competency that appear under the Situational Practice Domain to address specifically the CES motion regarding the TRC. The Situational Practice Domain comprises competencies that are to “focus on understanding, analyzing, and attending to the many circumstances that make every evaluation unique, including culture, stakeholders, and context.”

New: Uses evaluation processes and practices that support reconciliation and build stronger relationships among Indigenous and non-Indigenous peoples.

Previous: Shares evaluation expertise

Revised: Engages in reciprocal processes in which evaluation knowledge and expertise are shared between the evaluator and stakeholders to enhance evaluation capacity for all. (CES, n.d.)

The CES’s Essential Skills Series (ESS) is a series of entry-level courses developed by CES. The series is four days/20 hours in length and is intended to help participants gain basic levels of knowledge, skill, and appreciation with respect to the essential elements of evaluation. Upon completion, it is hoped that participants will be able to enter and participate in evaluation and develop as evaluation professionals. The Essential Skills Series has also undergone revisions. A Diversity Working Group Co-Chair, in partnership with the Mi’kmaw Native Friendship Centre, the Canadian Evaluation Society, and the Nova Scotia Health Research

Foundation used this as an opportunity to adapt the ESS to include Indigenous knowledge and culture. This initiative also included a day of reflection during which there was a discussion of which content and delivery methods could make ESS more culturally responsive. The feedback has been provided to the Board for incorporation into the next round of ESS revisions.

Another initiative that CES is supporting in its efforts in reconciliation is at the global level. It supports a Global Network called EvalIndigenous, of which I am the co-chair. EvalIndigenous is attempting to

Advance the recognition, value and involvement of Indigenous peoples in global evaluation practice and endeavors and,
Promote and Support Indigenous peoples' self-determination of their evaluation agenda. (Bremner & Were, 2016)

It is time for Indigenous peoples to determine their evaluation agendas. But how do we as evaluators support this endeavour? We need to let Indigenous peoples set and control their own evaluation and research agendas.

While our national organization has taken first steps in attempting to address reconciliation, what can we do as individual evaluators to incorporate reconciliation into our work?

EVALUATORS AS STORYTELLERS

I believe evaluators are storytellers. When you look at our world today, the following quotation by Ben Okri (1997, p. 46), a Nigerian storyteller, is more relevant than ever:

In a fractured age, when cynicism is god, here is a possible heresy: we live by stories, we also live in them. One way or another we are living the stories planted in us early or along the way, or we are also living the stories we planted—knowingly or unknowingly—in ourselves. We live stories that either give our lives meaning or negate it with meaninglessness. If we change the stories we live by, quite possibly we change our lives.

The importance of storytelling is well documented, and the important elements of culture, including language, customs, spirituality, history, and locality, all add to the stories we have to tell (Chouinard & Cousins, 2007; Kirkhart, LaFrance, & Nichols, 2011; Nee-Benham, 2008; Wilson, 2008). A recent article highlighted how Tribal Critical Theory (TCT) recognizes the importance of stories. As TCT suggests, not only are tribal beliefs, philosophies, and customs important for understanding the lived realities of Indigenous peoples, but “it [TCT] also recognizes the importance of story as a legitimate data source and building block of theory and insists that the interconnected nature of theory and practice demands that researchers work towards social change” (Bowman, Francis, & Tyndall, 2015, p. 338).

In many cases, we have not told the stories that need to be told. As storytellers we must be aware, when we approach our communities, of the trauma that is still being felt in many communities. As responsible evaluators we need our work to address cultural repression and support cultural renewal. We have to give back the knowledge we have been taking from the communities, and we have to help them become sovereign nations with the power and ability to control their own destinies. As noted by [Kirkhart et al. \(2011\)](#),

Historical trauma must be addressed. Communities need to heal from cultural repression, including repression for which research has been culpable. Evaluation must contribute to learning that supports cultural renewal and revitalization. Self-determination must be heard and understood by evaluators as a necessary condition of good evaluation.

THE STORIES WE USUALLY HEAR

Canadians routinely are presented with information concerning the lifespan of Indigenous adults, which we are told is 15 years shorter than for non-Indigenous adults ([CBC, 2018b](#)). We are informed that for Indigenous peoples, the incidence of diseases such as diabetes is four times the rate for non-Indigenous Canadians. Canadians are repeatedly told that there is an overrepresentation of Indigenous peoples when looking at infant mortality, suicide, and incarceration rates. These are the stories that are being told, which, according to [Allan and Smylie \(2015\)](#), are “frequently presented without the context needed to make sense of the information provided.” Furthermore, they go on to suggest not only is the context absent, but stories of Indigenous health in Canada also can be “characterized by the presence of racist stereotypes and inaccuracies pervasive in mainstream Canadian narratives.”

Another story with which we have become familiar is the number of Indigenous communities in Canada without running water and living under water advisories. As of August 2019, there were 56 long-term (i.e., longer than 12 months) advisories in effect ([Indigenous Services Canada, 2019](#)). These stories should “remind us that First Nations, Inuit and Métis peoples’ health status reflects the socioeconomic, environmental, and political contexts of their lives, a context inextricable from past and contemporary colonialism” ([Greenwood, de Leeuw, & Lindsay, 2018](#)).

Other stories we hear concern the experiences of Indigenous peoples in the justice system. In Canada there have been a number of high-profile justice-related cases involving Indigenous victims ([CBC, 2018a](#)). The results from these trials have led to questions regarding a justice system seen to be racist. A non-Indigenous male in Winnipeg was found not guilty of killing Tina Fontaine, a 15-year-old Indigenous female. Two weeks earlier, a non-Indigenous male in Saskatchewan, accused of killing Colten Boushie, a 22-year-old Indigenous man, was also found not guilty. These verdicts sparked outrage and were delivered almost a year after the Prime Minister of Canada called for a review of laws and

policies related to Indigenous peoples ([Government of Canada, 2017](#)). One must ask, as did Josée Lavoie, head of the Centre for Aboriginal Health Research at the University of Manitoba, “as a researcher, I wonder how many reports ... we need until we ask for action[?]” ([CBC, 2015](#)).

As evaluators, we know there have been many “reliable” and “valid” studies undertaken in Indigenous communities, and in many cases nothing has changed for those communities. Maybe it is time to think beyond reliability and validity and start talking about “authentic” work, work that is meaningful to the communities:

The challenge for future programs and their evaluations is to be part of the transformation and legitimization of knowledge ... issues of reliability and validity can be less important than the political and social realities that surround programs and communities wherein they operate. Equity and social justice should become the touchstones of practice, rather than the handmaidens of “reliability” and “validity.” Rigor gives way to authenticity. And regardless of these notions, credibility becomes the servant of audience. ([Lee & Bremner, 2012](#), p. 66)

Evaluation must build on the communities’ cultural, social, and spiritual values and support cultural resurgence. The focus of an Indigenous approach should not be on individuals and independence but on relationships and the community/collective. We must ask ourselves as evaluators the following questions: Is our work perpetuating contemporary colonialism? Are we doing things in a way that preserves the status quo? What can we do to help ensure that as Indigenous peoples we have the power to set and control our own evaluation and research agendas?

OUR ROLE IN TRUTH AND RECONCILIATION

When I interviewed a young Indigenous woman who is a leader in her community, we talked about reconciliation. She mentioned that in her community reconciliation is viewed as helping white people feel less guilty about themselves because, while there has been a lot of talk about reconciliation, nothing has changed in her community. A few months later, when talking to a colleague who had just spoken with an Elder, she mentioned that the Elder wanted to know what had happened to the “truth.” It was the Truth and Reconciliation Commission, and everyone just talks about reconciliation. The stories we tell and should tell as evaluators must deal with *both* truth *and* reconciliation. The historical and current-day impact of colonial policies speaks to the truth. Evaluators must understand the past realities of the communities in which we work and the roles we have played in perpetuating these truths. Through self-reflection and developing our understanding, evaluators will be able to move to reconciliation and a vision for the future in which Indigenous peoples control and develop their own evaluation agendas.

As mentioned earlier, there are 80,000 survivors of residential schools still living and there are stories that still have to be told; there are truths that have to be

heard. I believe our role as evaluators is to try to make sure not only that we hear those stories but also that these stories are heard by people in power, policymakers and decision makers in government. That is our responsibility in addressing the “truth” part of truth and reconciliation.

In terms of reconciliation, what can we do that will help change these stories so that visions for the future become reality? What can we do differently in our work? What can we change about how we conduct our enquiries? How can we move beyond sharing and return the power to the communities so we can change the larger story? These are the questions we have to ask ourselves. We have to start acting in reciprocal ways, so we are not only taking but also giving back to the communities in which we work.

EVALUATION AND HUMAN AND NATURAL SYSTEMS

A number of years ago, I was undertaking case studies in Nunavut. As I was talking with one individual, he made this point: “Rocks are living things. Rivers are living things. Ice is a living thing. If we respect them, they will give us what we need to live in harmony with the system We need to ensure a place for both world views.” Traditionally, evaluators have looked only at human systems. Recently, we have started to talk about evaluation in human and natural systems and how we need to start dealing with both systems in our work.

Evaluators frequently talk about silos, and we hear about silos in government; however, as evaluators we have been undertaking our work in silos. Funders, for the most part, have not funded holistic evaluations. They fund evaluations of individual initiatives—an education initiative or a health initiative or a justice initiative; however, in the communities in which I work, these are interconnected. Sometimes you have to address health and wellness before you can look at education, and sometimes you need to address justice issues before you deal with education. If evaluators continue to work in silos and if funders continue to fund in silos, we will never address some of the issues that need to be resolved. There is a need to move toward a more holistic approach, one that better reflects the relationality of Indigenous world views in which the human and natural systems are one system composed of relationships with the land, culture, community, people, ancestors, and spirituality.

Until recently I believed that Two-Eyed Seeing was a useful metaphor to take the Indigenous ways and meld them with the Western ways: “Two-Eyed Seeing—To see from one eye with the strengths of the Indigenous ways of knowing, and from the other eye with the strengths of the Western ways of knowing, and to use both of these eyes together” (Hatcher, Bartlett, Marshall, & Marshall, 2009, p. 3). My concern—and it is the same concern I have with co-creation and cultural competency—is that, while they are interesting concepts, they have the potential to maintain contemporary colonialism. My experience with Two-Eyed Seeing is that one eye is always bigger—and it is not the Indigenous eye. We have to give control back to Indigenous people

and their communities; they must be the ones who control their evaluation and research agendas.

MOVING FORWARD

The Indigenous peoples of Canada are not a homogeneous group. According to the 2016 Census, Indigenous peoples made up 4.9% of the total Canadian population. First Nations people, those who are registered or treaty Indians under the *Indian Act* and those who are not, make up 58% of the Indigenous population, followed by Métis (35%) and Inuit (4%) ([Statistics Canada 2017](#)).

Furthermore, there are many First Nations communities in Canada: more than 630, representing more than 50 First Nations, many of which have experienced different political and contextual realities. The Government of Canada estimates there are approximately 50 Indigenous languages in 12 language families, while UNESCO estimates that there are approximately 90 Indigenous languages. Our efforts in reconciliation will need to be as diverse as the populations with whom we are privileged to work.

I was discussing reconciliation with an individual I recently interviewed. We both noted that, in Canada, there have been many ongoing discussions regarding reconciliation at all levels. However, he made the point that this is just the beginning of our journey in reconciliation as a country. He stated that this journey will involve long, hard, difficult discussions; some people will get hurt and upset. However, these are the types of discussions we need as a country—and a profession—as we move forward.

How will the CES initiatives, such as the Essential Skills Series and the E-Institute, address the new competencies? We have a membership category for individuals from the Global South, which is a positive step. Are there changes that CES can make to its membership categories to attract Indigenous practitioners? In doing so, we may need to define evaluators in non-traditional ways; it might be an individual working in an Indigenous community who is a community development officer, it might be an Elder, or the principal of a school in an Indigenous community. These are individuals who are really important for us to welcome into our association and into our work.

These are the questions I am going to be asking myself continuously: How can we ensure that we incorporate truth and reconciliation into our work—my work—into what CES does as an organization? What stories need to be told and who will tell them? How will these stories speak to truth and reconciliation? How can we ensure that these stories will be heard?

As evaluators “we need to see Indigenous peoples as creators of their own destinies and experts in their own realities. ... [evaluation as] a source of enrichment to their lives and not a source of depletion or denigration” ([Weber-Pillwax, 1999, p. 38](#)). As Justice Murray Sinclair, Chair of the TRC reminds us, “The road we travel is equal in importance to the destination we seek. There are no shortcuts. When it comes to truth and reconciliation, we are all forced to go the distance” ([TRC, 2012, p. 1](#)).

REFERENCES

- Allan, B., & Smylie, J. (2015). *First peoples, second class treatment: The role of racism in the health and well-being of Indigenous peoples in Canada*. The Well Living House Action Research Centre for Indigenous Infant, Child, and Family Health and Wellbeing, St. Michael's Hospital. Retrieved from <http://www.wellesleyinstitute.com/wp-content/uploads/2015/02/Summary-First-Peoples-Second-Class-Treatment-Final.pdf>
- Bowman, N. R., Dodge Francis, C., & Tyndall, M. (2015). Culturally responsive Indigenous evaluation: A practical approach for evaluating indigenous projects in tribal reservation contexts. In S. Hood, R. Hopson, & H. Frierson (Eds.), *Continuing the journey to reposition culture and cultural content in evaluation theory* (pp. 335–359). Charlotte, NC: Information Age.
- Bremner, L. K., & Were, L. (2016). EvalIndigenous concept paper. Retrieved from <http://forum.ioce.net/forum/open-forums/evalindigenous>
- Canadian Broadcasting Corporation (CBC). (2015). *First nations adults more than twice as likely to die from avoidable causes*. Retrieved from <https://www.cbc.ca/news/indigenous/first-nations-adults-more-than-twice-as-likely-to-die-from-avoidable-causes-1.3196496>
- Canadian Broadcasting Corporation (CBC). (2018a). *High-profile court cases highlight difficulties for Indigenous youth, say leaders*. Retrieved from <https://www.cbc.ca/news/canada/saskatoon/colten-boushie-tina-fontaine-verdicts-indigenous-youth-1.4549049>
- Canadian Broadcasting Corporation (CBC). (2018b). *Lifespan of Indigenous people 15 years shorter than that of other Canadians, federal documents say*. Retrieved from <https://www.cbc.ca/news/health/indigenous-people-live-15-years-less-philpott-briefing-1.4500307>
- Canadian Evaluation Society (CES). (2016a, May 17). The CES takes position regarding reconciliation in Canada. Retrieved from <https://evaluationcanada.ca/news/6348>
- Canadian Evaluation Society (CES). (2016b, June 21). Letter to Carolyn Bennett. Retrieved from https://evaluationcanada.ca/sites/default/files/bennett_carolyn_20160621_en.png
- Canadian Evaluation Society (CES). (n.d.). Competencies for Canadian evaluators. Retrieved from <https://evaluationcanada.ca/competencies-canadian-evaluators>
- Chouinard, J., & Cousins, B. (2007). Culturally competent evaluation for aboriginal communities: A review of the empirical literature. *Journal of MultiDisciplinary Evaluation*, 4(8), 40–57.
- Government of Canada. (2017). *Prime minister announces Working Group of Ministers on the Review of Laws and Policies Related to Indigenous Peoples*. Retrieved from <https://pm.gc.ca/eng/news/2017/02/22/prime-minister-announces-working-group-ministers-review-laws-and-policies-related>
- Greenwood, M., de Leeuw, S., & Lindsay, N. (2018). Challenges in health equity for Indigenous peoples in Canada. *The Lancet*, 391(10131), 1645–1648. Retrieved from [https://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(18\)30177-6/fulltext](https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(18)30177-6/fulltext)
- Hatcher, A., Bartlett, C., Marshall, M., & Marshall, A. (2009). Two-eyed seeing: A cross-cultural science journey. *Green Teacher*, 86, 3–9.

- Indigenous and Northern Affairs Canada. (n.d.). Indian Residential Schools. Retrieved from <https://www.aadnc-aandc.gc.ca/eng/1100100015576/1100100015577#sect1>
- Indigenousfoundations.arts.ubc.ca. (2009). University of British Columbia. Retrieved from https://indigenousfoundations.arts.ubc.ca/the_residential_school_system/
- Indigenous Services Canada. (2019). Ending long-term drinking water advisories. Retrieved from <https://www.sac-isc.gc.ca/eng/1506514143353/1533317130660>
- Kirkhart, K., LaFrance, J., & Nichols, R. (2011). *Improving Indian education through Indigenous evaluation*. American Educational Research Association Conference, New Orleans, LA.
- Lee, L., & Bremner, L. (2012). Evaluation frameworks and educational equity: Friends or foes? In D. Pop (Ed.), *Education policy and equal education opportunities* (pp. 51–68). New York, NY: Open Society Foundations.
- Legacy of Hope Foundation. (2014). *Forgotten: The Métis residential school experience*. Ottawa, ON: Author.
- Nee-Benham, M. K. P. A. (Ed.). (2008). *Indigenous educational models for contemporary practice: In our mother's voice* (Vol. II). New York, NY: Routledge, Taylor and Francis.
- Okri, B. (1997). *A way of being free*. London: Weidenfeld & Nicolson/Phoenix House.
- Statistics Canada. (2017). *Aboriginal peoples highlight tables, 2016 Census*. Retrieved from <https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/hltfst/abo-aut/Table.cfm?Lang=Eng&T=101&S=99&O=A>
- Truth and Reconciliation Commission of Canada (TRC). (2012). *Interim Report*. Retrieved from http://www.trc.ca/assets/pdf/mroom_Interim%20report%20English%20electronic.pdf
- Truth and Reconciliation Commission of Canada (TRC). (2015). *Honouring the truth, reconciling for the future: Final report of the Truth and Reconciliation Commission of Canada*. Retrieved from http://www.trc.ca/assets/pdf/Honouring_the_Truth_Reconciling_for_the_Future_July_23_2015.pdf
- Weber-Pillwax, C. (1999). Indigenous research methodology: Exploratory discussion of an elusive subject. *Journal of Educational Thought*, 33(1), 31–45.
- Wilson, S. (2008). *Research is ceremony: Indigenous research methods*. Winnipeg, MB: Fernwood.

AUTHOR INFORMATION

Larry Bremner is a former CES national president, CES Fellow, and award winner. He established Proactive Information Services Inc. in 1984 to provide evaluation services to the not-for-profit and public sectors. He has worked across Canada in urban, rural, remote, and Indigenous communities, and throughout Europe. Larry was the driving force behind the creation of EvalPartners' global network EvalIndigenous and was its first chair. Proud of his Métis heritage, he is passionate about equity and access. Larry believes we are compelled to create a future that is inclusive, if we are to address today's crucial social, environmental, and economic issues.

Nation-to-Nation Evaluation: Governance, Tribal Sovereignty, and Systems Thinking through Culturally Responsive Indigenous Evaluations

Nicole Bowman (Wapalaneexkweew)

University of Wisconsin – Madison / Bowman Performance Consulting

Abstract: *This paper was presented as part of the opening plenary panel at the 2018 Canadian Evaluation Conference in Calgary, Alberta, on May 27, 2018. Through telling the origin stories of First Nations/Indigenous people and Western evaluation colleagues, we can begin to understand the history and practical applications for advancing the truth through evaluation. The Doctrine of Discovery is rarely told as part of the Western canon of history or contemporary evaluation practice. There are significant and negative cultural, human rights, and social impacts that have deep institutional and systemic roots that continue to cause harm to First Nations/Indigenous populations throughout the world. To change centuries of old negative outcomes and impacts, we must understand our personal origin stories and the origin stories embedded within evaluation. Governance, policy, and evaluation can work as transformative levers for professional and sustained change if systems, critical and Indigenous theories, and methods are utilized. This paper offers origin stories of First Nations and colonial nations as a historical perspective and a new Tribal Critical Systems Theory to change contemporary Nation-to-Nation evaluation practices.*

Keywords: *culturally responsive evaluation, First Nations, government evaluation, Indigenous evaluation, Native American, sovereignty, systems evaluation*

Résumé : *Le présent article a été présenté dans le cadre de la séance plénière d'ouverture du Congrès annuel de la SCÉ à Calgary (Alberta) le 27 mai 2018. En racontant les récits des origines des Premières Nations/peuples autochtones et de nos collègues occidentaux en évaluation, nous pouvons commencer à comprendre l'histoire et les applications pratiques de faire connaître la vérité par l'intermédiaire de l'évaluation. La doctrine de découverte est rarement racontée dans le cadre des canons de l'histoire occidentale ou de la pratique contemporaine de l'évaluation. Des effets significatifs et négatifs en matière de culture, de droits de la personne et de société, qui ont des racines institutionnelles et systémiques profondes, continuent de causer du tort aux Premières Nations/populations autochtones partout dans le monde. Pour changer des siècles de résultats et d'impacts négatifs, nous devons comprendre notre origine personnelle et les histoires d'origine de l'évaluation. La*

Corresponding author: Nicole Bowman (Wapalaneexkweew), Wisconsin Evaluation Collaborative (WEC) Center, University of Wisconsin – Madison, 1025 West Johnson Street, Suite 1152, Madison, WI, 53706 USA; nrbowmanfarr@wisc.edu or nicky@bpcwi.com

gouvernance, les politiques et l'évaluation peuvent servir de levier de transformation pour les professionnels et pour le changement durable si les théories et les méthodes autochtones sont utilisées. Cet article présente des histoires sur l'origine des Premières Nations et des Nations coloniales dans une perspective historique, ainsi qu'une nouvelle théorie tribale des systèmes critiques pour faire évoluer les pratiques d'évaluation contemporaine de Nation à Nation.

Mots clé : *évaluation adaptée à la culture, Premières Nations, évaluation gouvernementale, évaluation autochtone, Amérindien, souveraineté, évaluation des systèmes*

[This paper was presented as part of the opening plenary panel at the 2018 Canadian Evaluation Conference in Calgary, Alberta, on May 27, 2018. Through telling the origin stories of First Nations/Indigenous people and Western evaluation colleagues, we can begin to understand the history and practical applications for advancing truth through evaluation. Both verbatim excerpts from Dr. Bowman's presentation as well as additional information from the original keynote paper are provided in this article. Additional information is presented in this article, and it expands upon the original keynote address, which was limited due to the time constraints. Dr. Bowman's original keynote address was one presentation within a panel of four presenters, each limited to a 10–12 minute presentation. For the original keynote addresses, see [Bowman, Bremner, McKegg, and Wehi-peihana \(2018\)](#).

Dr. Bowman's opening plenary remarks began with her introducing herself: "Good Morning! Hello! We're sold out, we should be excited! My Indigenous spirit name, given to me by my Mohican and Lunaape-Munsee (e.g., Delaware) traditional leaders, is Lenapexkwe Waapalaneexkweew Neeka Ha Newetkaski Nawa Oplanwuuk. This translates roughly to Flying Eagle Woman Who Is Accompanied by Four Eagles. My clan is Laaweewapooosish (Lynx) and I reside in Wisconsin as part of the Stockbridge-Munsee Nation in Bowler, Wisconsin, USA. I would like to say Anushiik (thank you) to all the conference folks who believed in me. When the Canadian Evaluation Society (CES) asked me to be a plenary speaker, the first thing they said to me was, "you don't behave, that's why we're picking you!" My response was, "YES!" As I said, I am from the Stockbridge-Munsee Community. I'm Mohican and Lunapee and my mother is Polish. I identify culturally and practically first as a traditional Mohican woman, bi-racial, and a knuckle-dragging, blue-collar scholar. I'm going to talk a little bit today about going beyond feathers and beads, and I don't mean that in a disrespectful way. I've done a lot of work to prepare to come here; my Elders have prepared me through our ceremonial and traditional ways. I'd like to recognize my Elders in the room, and my ancestors who spiritually may be floating around too. And in my community, calling someone an Elder, with a capital "E," is a sign of respect. I would just like to say that I'm proud to be living my traditions through my academic work and this plenary. And it is from this traditional ground from which I stand

on, moccasins and all, that I share my plenary perspectives with you today. Please know that I am very humbled and honored to be chosen to be here today. I thank the Indigenous ancestors of whose bones I stand on. This sacred responsibility is not lost by me.”]

SPEAKING YOUR TRUTH THROUGH POWERFUL CREATION STORIES

“Speaking truth to power opened my eyes to our job as evaluators and how much the truth influences those engaged in all aspects of the evaluation. Truth is a powerful tool. Speaking truth to power made me realize that honesty is one of my greatest gifts. Providing the truth is one of my greatest powers.”

—D. Pingel (Brothertown Nation)

The 2018 Canadian Evaluation Society (CES) conference theme of “co-creation” was about meaningfully engaging people to create mutually valued outcomes together, collaborating towards common goals. The foundation to making good relations with First/Tribal Nations and Indigenous people must be founded on truth. Shaaxkaaptóoneew (speaking the truth, straight talk) is available to everyone, but few have the courage, humility, or discipline to ground themselves in this sacred manner. Truths regarding First/Tribal Nations are uncomfortable and traumatic. However, they are supported by historical and contemporary facts—facts which are often unknown or ignored by western (i.e., colonized) institutions, systems, and people for various resource, capacity, or ethical reasons (Waapalaneexkweew & Dodge-Francis, 2018). Nevertheless, in instances where Indigenous and non-Indigenous people responsively, respectfully, and resourcefully work together, the truth provides a fertile ground for transformative and sustainable change.

For example, the Lunapee are also called Delaware Indians; the colonized words are Delaware and Indians. Those labeled as Delaware Indians would identify themselves as Lunaape and Mohican Nations. This is important to know because pre- and early contact with European/colonial populations included working under treaties that recognized Lunaape (and other Native/First Nations) as sovereign Nations. Work between Nations still should be governed by treaties, for First Nations and Canada as well as Tribal Nations and the United States. All have treaty law ensconced in contemporary constitutional laws—starting first in the Eastern door where the sun rises and the color is red for Native/First Nations, talking about the origin stories of Lunaape people and Mohicans coming from Kukakha Ahkuy (Mother Earth). Lunaape call this place Turtle Island, which is now called the United States, Canada, and North America. This becomes important in the Lunaape creation story. It is illustrated within the decolonized map in [Figure 1](#). How Lunaape define and talk about who they are and where they are now as Indigenous people living in North America really shows the level of assimilation and

colonization (Wapalaneexkweew, 2018, p. 5). This is important to understand the Lunaape origin story and the western/colonizer origin story via the Doctrine of Discovery, which will be discussed later.

The responsibility is ours (i.e., the field of evaluation), not just mine or that of Indigenous people and First Nations governments. Knowing how history happened from diverse and accurate perspectives, how western/colonizer history impacts Lunaape history, and explicitly linking those histories to the contemporary legal, policy, governance, and sovereignty aspects that should be embedded in evaluation are requirements for all. Evaluators are standing on contested land. Evaluators make designs, collect data, and offer recommendations about Nations and people. Therefore, evaluators have a professional, ethical, and moral obligation to seek knowledge about and understanding of the Nations and people where they work. These responsibilities speak not only to personal scholarship but also to personal, cultural, and academic competencies, humility, and capacities as a professional evaluator. Typical Indigenous approaches include smudging, singing or nods to culture, and language, but evaluators must move to the legal, political, and governance systems aspects of evaluation with Indigenous people and First Nations. This is where the rubber hits the road, or where theory becomes practice. Without legal and political frameworks and theories to inform policy and evaluation, there will be no institutional or systemic changes. Therefore, it is important to understand origin stories, and how these apply to the field of evaluation and individual professional evaluation practice.

Figure 1. Mohican Nation: The Munsee/Lunaape (Delaware) Trail of Tears Map

THE UNVARNISHED TRUTH: WESTERN CREATION STORIES BEGIN WITH THE DOCTRINE OF DISCOVERY

Evaluators ask questions, investigate, apply tools, collect data, and measure results. Evaluators combine expertise and objectivity to understand different perspectives, realities, and interests. Evaluators often operate from a position of privilege and influence, representing a funder or the agency responsible for delivering a program. Traditionally, clients or target groups are perceived as object(s) of the evaluation. Is it possible that evaluators ignorantly privilege their personal worldviews or methodological choices? Do evaluators carry in epistemic or other privileges (Fricker, 2007; Grosfoguel, 2013)? Or do evaluators truly make space to honor and incorporate the knowledge, values, culture, and lived experiences of those who engage with us, thus resulting in an evaluation that is by and for Indigenous people and Tribal/First Nations (Prussing, 2018; Rodriguez-Lonebear, 2016; Wehipeihana & McKegg, 2018)? Developmental evaluation, participatory evaluation, and empowerment evaluation have attempted to address power imbalances through a variety of methods (Patton, McKegg, & Wehipeihana, 2015). Indigenous and decolonized framing of evaluation (Bowman, Dodge-Francis, & Tyndall, 2015; Commonwealth of Australia, 2018; LaFrance & Nichols, 2008) have also focused on addressing power imbalances. However, adequate and sustainable resources for collaboration and reciprocal relations often fall short of true partnerships with First/Tribal Nations and community members. Applying the 2018 Canadian Evaluation Society's conference theme of co-creation, this article will unpack the concept of evaluation without dominion. In the Canadian context (and in other colonized nation-states where Indigenous people were conquered), the language of dominion matters. When Canada, or other nation states, conquered Indigenous people, the destruction and political or Christian justification to do so is documented through the Doctrine of Discovery (Alexander VI, 1493; Cobb, 2015). Dominion is about territory at the time of the earliest contact of Europeans with North America. In the field of evaluation, it is primarily people of European descent that control the "original" evaluation narratives and are often most cited in theory, method, and practice. Unpacking the complexities of evaluation must be done with unconquered minds. Bringing an awareness or open-mindedness to learning about concepts of Indigenous evaluation (original evaluation) needs to be done without dominion. To do this requires starting from the beginning. What is the Western/colonizer "creation story" and that of the institutions and systems from where evaluation begins? How is the Western/colonizer "creation story" informed by policy and funding? If it is unknown, then trying to use evaluation as a tool for emancipatory and empowering practice to solve social and other issues will be a fleeting effort. Worse than that, evaluators are likely to be replicating trauma and harm, whether or not they realize it.

The creation story for Western civilization and institutions begins with the Doctrine of Discovery (DD). The DD is the religious and legal justification for the domination and destruction of Indigenous populations (Cobb, 2015). The DD was written in 1493 (Alexander VI, 1493) and has been carried out for centuries

thereafter. This Conqueror model (Newcomb, 2008) continues to have contemporary policy, practice, and academic impacts, including impacts in evaluation (Waapalaneexkweew, 2018). It is important that evaluators and other non-Indigenous people have an understanding of the Western civilization origin story and the DD, as it relates to the context(s) in which evaluators work. They should know how the history of the land and Western/colonizer history is explicitly and legally linked to policy, governance, and sovereignty of the contested land that everyone is standing on. This is all very important for the field of evaluation and for evaluators' academic and professional evaluation practices. In addition to incorporating cultural protocols for Indigenous populations (e.g., smudging, culture, language, etc.), evaluators must also consider the legal, political, and policy aspects of sovereign Tribal/First Nations when conducting evaluations.

The Truth and Reconciliation Commission (TRC) in Canada includes 94 recommendations, of which 44 have been started or completed, according to the Beyond 94 website (CBC, 2018). The remaining recommendations have not been started or are in the very early stages of implementation. These slower-to-implement recommendations (and difficult-to-evaluate impacts and changes) are mostly about the legal and policy changes needed for sustainable impacts regarding the TRC and Indigenous people and Tribal/First Nations. Simply put, the education, outreach, and public service announcements regarding the TRC are not enough. The hard work will be realized when policy, practice, and evidence of change are being documented through responsive evaluation. These legal and policy foundations are critical to the short- and long-term success of the TRC. This authentic work must include and allow Tribal/First Nations to lead or co-lead the efforts. A "Nation-to-Nation Agenda" is what is needed in order to realize that "indigenous rights do not derive from the [Canadian] Constitution" (Shepherd & McCurry, 2018). Therefore, it is critically important to understand what happened, how long ago it happened, how it applies to individuals, and how it applies to individuals' evaluation and policy work. This is very important, even though it may seem that the DD happened long, long ago.

In the United States, the Marshall Trilogy, written in the early nineteenth century, demonstrates the "Constitutional dehumanization of American Indians" (Goetting, 2010). These are the case laws that are situated within our contemporary constitution. They continue to undermine Tribal sovereignty and harm Tribal Nations and communities to this very day in the United States (National Congress of American Indians, 2018; Parker, 2018). Hence, non-Indigenous stakeholders need to know the whole story. The story behind protests of oil pipelines, diamond, and/or iron ore mines. The story behind unaddressed issues of missing and murdered Indigenous women, forced removals of children (e.g., boarding or residential schools, foster care system, etc.), and related intergenerational trauma. The story behind the impacts of assimilationist policies that continue to come up in policy, practice, and academic studies. Again, to reiterate, non-Indigenous stakeholders need to know the whole story, from the beginning, because not much has changed, including the devastating and disproportionate educational,

social, economic, and health outcomes for Indigenous people. Oral history is not recognized as legitimate evidence under the US constitutional government. The academics doing the research and making the rules about what evidenced-based policy and programs are practicing (Bipartisan Policy Center, 2018; Schoenefeld & Jordan, 2017) continue to omit Native/First Nations people from history and contemporary participation in broad policy, leadership, and academic initiatives. Canada and the Canadian Constitution are not much different when it comes to policy, governance, and evaluation practices with First Nations (Canadian Press, 2018; Diabo, 2018; Green & Starblanket, 2018). For Canada to fully embrace the *Principles Respecting the Government of Canada's Relationship with Indigenous Peoples* (Government of Canada, Department of Justice, 2018) or implement the *United Nations Declaration on the Rights of Indigenous Peoples* (UNDRIP) (United Nations, 2008), a change in theory, method, and practice will be needed at all levels (i.e., systems, institutional, and practitioner). Evaluators can be facilitators to the critical levers of change that are needed to make policies, principles, and declarations operational and impactful. Through research on evaluation, evaluation policy, evaluation technical assistance, and development activities, and as co-leaders with Indigenous scholars and Tribal/First Nations designing and implementing evaluations, the field of evaluation can be the catalyst for needed changes that should have happened centuries ago.

CRITICAL TRUTH TELLING: AN EMERGING TRIBAL CRITICAL SYSTEMS THEORY

Co-creation of an evaluation and a re-thinking of our evaluative practice challenge traditional power relationships. It requires an evaluator to be a methodological expert, facilitator, critic, ally, and strategic thinker with the ability to move evaluation toward empowering change, while sharing multi-jurisdictions (Reinhardt & Maday, 2005), Tribal/First Nations governments, other public governments, and non-profit or private-sector partners need to collaboratively work together on initiatives. It requires knowledge of origin stories, developing trusting and reciprocal partnerships, properly resourcing evaluations from the ground up, and acknowledging that other methods and leadership/governance perspectives (i.e., sovereignty) are equally influential to non-Indigenous governments and agencies. Using a government-to-government framing of evaluation, it is proposed that the field of evaluation should consider how an emerging Tribal Critical Systems Theory (TCST) can be applied to evaluation. Blending systems theory and thinking, critical systems theory, Tribal Critical Theory (TCT), and Indigenous Evaluation (IE) can begin to conceptualize how Tribal sovereignty can be raised to a systems level. At a systems level it can influence evaluation policy and evidence-based practice through Tribal/First Nations and public government initiatives.

TCST provides an emerging framework for raising critical consciousness, activating culturally responsive practices in systems, and emancipating public government, policy, and practices. TCST reflects more equitable and valued

approaches to working with (not on) Tribal/First Nations government partners. Indigenization can be a tool for inclusion, reconciliation, and decolonization and can also be a new vision for how Canadian academics respond to the TRC's calls to action (Gaudry & Lorenz, 2018). Indigenous visioning can be applied not only to the TRC's calls to action but also to systems evaluation or other evaluation efforts with Tribal/First Nations governments and Indigenous people globally. Table 1 offers a way to begin thinking through a new TCST by building on Tribal Critical Theory (Brayboy, 2005). TCST was developed through evaluation policy and practice applications to academic and professional activities that I implemented in the US geographic context. This emerging theory could be tested and applied to systems thinking for Nation-to-Nation activities as Tribal and non-Tribal governments work together on policy, research, and evaluation across North America or in other Indigenous contexts and evaluation initiatives globally.

Table 1. Applying TCT to systems for an emerging TCST Nation-to-Nation evaluation model

Tribal Critical Theory (Brayboy, 2005)	Toward a new Tribal Critical Systems Theory (TCST) (Bowman)
Colonization is endemic to society	The political power of public governments was achieved and is sustained through illegal, unjust, and unethical means
US policies toward Indigenous peoples are rooted in imperialism, White supremacy, and a desire for material gain	Public government constitutions and policies are founded on Christianity and the Doctrine of Discovery, both of which did not consider Indigenous people humans, and continue currently through directly related federal and case law
Indigenous peoples occupy a liminal space that accounts for both the political and racialized natures of our identities	Tribal/First Nations governments and people are the only racial/ethnic group that has inherent political and legal rights equal to US federal and international governments through treaties and constitutional law
Indigenous peoples have a desire to obtain and forge tribal sovereignty, tribal autonomy, self-determination, and self-identification	By UN Resolution, the UNDRIP (2008) provides 46 articles that outline the global rights of Indigenous people and Tribal/First Nations, which most countries have formally agreed to
The concepts of culture, knowledge, and power take on a new meaning when examined through an Indigenous lens	Culture, knowledge, and power are defined uniquely and locally through both traditional Indigenous governments and contemporary Tribal/First Nations constitutions, ordinances, policies, and community practices

Tribal Critical Theory (Brayboy, 2005)

Governmental and educational policies toward Indigenous peoples are intimately linked around the problematic goal of assimilation

Tribal philosophies, beliefs, customs, traditions, and visions for the future are central to understanding the lived realities of Indigenous peoples but also illustrate the differences and adaptability among individuals and groups

Stories are not separate from theory; they make up theory and are, therefore, real and legitimate sources of data and ways of being

Theory and practice are connected in deep and explicit ways such that scholars must work toward social change

Toward a new Tribal Critical Systems Theory (TCST) (Bowman)

Tribal/First Nations government and educational policies are strengths-based, locally define, and have equal (or more) influence in federal (or state and municipal) public policy and non-Tribal federal (or state and municipal) government contexts

Traditional, cultural, and community-based philosophies, knowledge, and practices are the foundation to contemporary Tribal/First Nations governments who are working with non-Tribal governments to create responsive, effective, and sustainable systems, institutional, and policy changes

Traditional history and knowledge that are orally transferred are essential to the scholarly and culturally responsive development and implementation of more effective policies, programs, and models

Evaluating, generating, and replicating more effective Tribal/non-Tribal governance models that theoretically and practically provide better supports, improvements, and outcomes for sustained positive changes in Tribal/First Nations and Indigenous communities represent a professional and ethical responsibility for all government and academic partners

CONCLUDING THOUGHTS: THE TRUTH CAN SET EVALUATORS FREE

Today, evaluators are asked to think about their personal origin stories and that of the geographic place where living and work occur. Start there. Put in effort toward education from sources, places, and people that normally are not on standard reading lists and are not in the bibliographies of the most current published articles. Without new sources to interrupt and challenge current cognitive frameworks and lived experience, there is a risk of repeating what Columbus and the rest of the “discoverers” of the earth did. Understanding political and scientific colonialism becomes very important to the work of evaluators, policymakers, and government leaders. When evaluators engage in research or evaluation, then publish an article, was permission granted from the elders or entity with the authority to bestow blessings or grant such permissions? Is that in current policies? Is that in

current contract language? Are institutions hiring Indigenous scholars or sharing (or giving) intellectual and human-subject protections to Tribal/First Nations? If evaluators are doing studies and teaching others, is that a sacred responsibility? If not, evaluators should be passing it by traditional councils or Tribal/First Nations governments. Lunaape know what their sacred responsibility is because the Elders have taught them and hold them accountable to that on a daily basis. Why do some evaluators feel entitled to claim access and ownership to data, knowledge, findings, etc.? Traditionally, Lunaape believe that people do not own knowledge but are caretakers of it. Therefore, having cultural and academic humility and giving back to the Indigenous community are also Lunaape responsibilities. How and to whom are evaluators responsible?

Beyond considering the importance of not continuing colonial practices, as it relates to evaluation or other disciplines, what I am presenting here is an emerging way to think about the human, legal, and political rights of Tribal/First Nations and Indigenous people through TCST. I am also highlighting the growing literature and leadership on evaluation, including within Tribal/First Nations. Many of the references in this article offer future reading and ways to become engaged practically and professionally. If you identify as a non-Indigenous professional, please observe, experience, and listen—first and foremost. This practice is not intended to shame; it is focused on healing. Evaluation can be good medicine if done properly. Evaluators need to first understand their personal history and how that history sustained or destroyed others. From those truthful origin stories, everyone can begin not only to heal, but also to change present practice and future history.

In conclusion, the field of evaluation and evaluators must understand the origins of authority structures, power sources (e.g., political, financial, etc.), and systems. These support, create gaps, or continue to silence Tribal/First Nations and Indigenous voices. As Tribal/First Nations and Indigenous communities continue working with public governments and other non-Indigenous partners, the new TCST framing can help academic and government agencies and partners to reconceptualize how the field of evaluation might move forward. This re-framing of and re-commitment to the scholarship, humanization, and professionalization of the field of evaluation can be carried out through evaluation policy, evaluation research, and evaluation studies. Comprehensively, multiple strategies and multi-jurisdictional partnerships that are adequately resourced will be needed to generate future evidence-based policies, practices, and effective relationships within and across evaluation systems and institutional partners. It's time to break the sounds of silence. Will you join us? Together all evaluators can break the sounds of silence (Bowman, 2018).

(Break the) Sounds of Silence (video played at the CES 2018 Plenary; link at <https://www.youtube.com/watch?reload=9&v=7fpo5Lc1-U0>)

By: N. Bowman as adapted from Simon (1964)

Koolamalsi, darkness my old friend
I've come to talk with you again

Because the visions that I am seeing
 Are born from places when I am dreaming
 And the sunrise visions are still planted in my brain
 Ancestors teachings still remain
 Within the sound of Silence
 In your visions and dreams you never walk alone
 Your Ancestors and relatives are all around
 Our spirit, blood and bones are all within this land
 What has been done to us was part of the Christian and colonial plans
 And my eyes can't stand seeing these things in my spiritual sight
 We have to make it right
 Let's break the sound of silence
 And in the sunrise ceremony I saw
 Millions of traditional people, maybe more
 Ancestors are talking, we are listening!
 Our spirits are strong, we are rising!
 We are dancing and singing songs that our people have always shared
 We're not scared for we will break the sound of silence
 Fools who think that we do not know
 We will not let your silent cancer grow
 Hear our prayers so we may teach you
 Take our hands so we may reach you
 Let our words be like medicine raindrops that fall and stop this living hell
 Put your voices in the sound of silence
 Don't go back to bow and pray to the false gods and scholars that are made
 Our Ancestors prophesized this warning:
 The Seventh Generation is here, are you coming?
And the signs of our Ancestors and Warriors are written on Spiritual walls
 Can you hear them call?
 They are whispering to you in the sounds of Silence.

REFERENCES

- Alexander VI. (1493). *Inter caetera* [Papal encyclical]. Retrieved from <http://www.papalencyclicals.net/Alex06/alex06inter.htm>
- Bipartisan Policy Center. (2018, April 11). *Evidence use in Congress*. Retrieved from <https://bipartisanpolicy.org/library/evidence-use-in-congress/>
- Bowman, N. R. (2018). *Break the sounds of silence* [video]. Retrieved from <https://youtu.be/7fpo5Lc1-U0>
- Bowman, N. R., Bremner, L., McKegg, K., & Wehipeihana, N. (2018, May 27). *Reconciliation and culturally responsive evaluation: Rhetoric or reality?* Canadian Evaluation Society: CES Co-Creation 2018, Calgary, Alberta. Retrieved from <https://www.youtube.com/watch?v=B6AnckEH1Eg>

- Bowman, N. R., Dodge-Francis, C., & Tyndall, M. (2015). Culturally responsive Indigenous evaluation: A practical approach for evaluating Indigenous projects in tribal reservation contexts. In S. Hood, R. Hopson, & H. Frierson (Eds.), *Continuing the journey to reposition culture and cultural context in evaluation theory and practice* (pp. 335–361). Greenwich, CT: Information Age Publishing.
- Brayboy, B. M. J. (2005). Toward a Tribal Critical Race Theory in education. *The Urban Review*, 37(5), 425–446. <https://doi.org/10.1007/s11256-005-0018-y>
- Canadian Broadcasting Corporation (CBC). (2018, March 19). *Beyond 94: Truth and reconciliation in Canada*. Retrieved from <http://cbc.ca/beyond94>
- Canadian Press. (2018, November 22). TRC heard concerns about coerced sterilization of Indigenous women, says Murray Sinclair. CBC. Retrieved from <https://www.cbc.ca/news/indigenous/trc-concerns-coerced-sterilizations-murray-sinclair-1.4917321>
- Cobb, D. M. (Ed.). (2015). *Say we are nations: Documents of politics and protest in Indigenous America since 1887*. Chapel Hill, NC: University of North Carolina Press.
- Commonwealth of Australia, Department of the Prime Minister and Cabinet. (2018). *Indigenous advancement strategy evaluation framework*. Retrieved from <https://pmc.gov.au/sites/default/files/publications/ias-evaluation-framework.pdf>
- Diabo, R. (2018). Trudeau gov't's plan to entrench 4th level of Indigenous gov't and permanent subjugation of First Nations. *First Nations Strategic Bulletin*. Media Co-op. Retrieved from <http://www.mediacoop.ca/story/trudeau-government-%E2%80%9C9Crightrights-recognition-framework%E2%80%9D36775>
- Fricker, M. (2007). *Epistemic injustice: Power and the ethics of knowing*. New York, NY: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780198237907.001.0001>
- Gaudry, A., & Lorenz, D. (2018). Indigenization as inclusion, reconciliation, and decolonization: Navigating the different visions for indigenizing the Canadian Academy. *AlterNative: An International Journal of Indigenous Peoples*, 14(3), 218–227. <https://doi.org/10.1177/1177180118785382>
- Goetting, N. J. (2010). The Marshall trilogy and the constitutional dehumanization of American Indians. *Indigenous Policy Journal*, 21(3), 1–17. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1840824
- Government of Canada, Department of Justice. (2018). *Principles respecting the government of Canada's relationship with Indigenous peoples*. Retrieved from <https://www.justice.gc.ca/eng/csj-sjc/principles-principes.html>
- Green, J., & Starblanket, G. (2018, August 7). Recognition of Rights Framework risks terminating rights for Indigenous peoples. *Regina Leader-Post*. Retrieved from <https://leaderpost.com/opinion/columnists/recognition-of-rights-framework-risks-terminating-rights-for-indigenous-people>
- Grosfoguel, R. (2013). The structure of knowledge in Westernized Universities: Epistemic racism/sexism and the four genocides/epistemicides of the long 16th century. *Human Architecture: Journal of the Sociology of Self-knowledge*, 11(1), 73–90. Retrieved from <http://scholarworks.umb.edu/humanarchitecture/vol11/iss1/8>
- LaFrance, J., & Nichols, R. (2008). Reframing evaluation: Defining an Indigenous evaluation framework. *The Canadian Journal of Program Evaluation*, 23(2), 13–31.

- National Congress of American Indians. (2018, December 3). *Joint statement on the United States Department of Justice's decision to defend the Indian Child Welfare Act*. Retrieved from <http://www.ncai.org/news/articles/2018/12/03/joint-statement-on-the-united-states-department-of-justice-s-decision-to-defend-the-indian-child-welfare-act>
- Newcomb, S. (2008). *Pagans in the promised land: Decoding the doctrine of Christian discovery*. Golden, CO: Fulcrum.
- Parker, C. (2018, May 7). Trump's political maneuvers could rob Native people of Tribal sovereignty. *IC Magazine*. Retrieved from <https://truthout.org/articles/trump-s-political-maneuvers-could-rob-native-america-of-tribal-sovereignty/>
- Patton, M. Q., McKegg, K., & Wehipeihana, N. (Eds.). (2015). *Developmental evaluation exemplars: Principles in practice*. New York, NY: Guildford.
- Prussing, E. (2018). Critical epidemiology in action: Research for and by Indigenous peoples. *SSM—Population Health*, 6, 98–106. <https://doi.org/10.1016/j.ssmph.2018.09.003>. Medline:30246140
- Reinhardt, M., & Maday, T. (2005). *Interdisciplinary manual for American Indian inclusion* (1st ed.). Tempe, AZ: Educational Options, Inc. Retrieved from https://tribaled-departmentsna.files.wordpress.com/2016/08/american-indian_inclusionmanual.pdf
- Rodriguez-Lonebear, D. (2016). Building a data revolution in Indian country. In T. Kukutai & J. Taylor (Eds.), *Indigenous data sovereignty: Toward an agenda* (pp. 253–272). Canberra, Australia: Australian National University Press.
- Schoenefeld, J., & Jordan, A. (2017). Governing policy evaluation? Towards a new typology. *Evaluation*, 23(3), 274–293. <https://doi.org/10.1177/1356389017715366>
- Shepherd, R., & McCurry, P. (2018, October 31). Ottawa must talk to Canadians about nation-to-nation agenda: It's time for the Trudeau government to fully engage Canadians and Indigenous communities in the conversation around its nation-to-nation agenda. *Policy Options Politiques: The Public Forum for the Public Good*. Retrieved from <http://policyoptions.irpp.org/magazines/october-2018/ottawa-must-talk-to-canadians-about-nation-to-nation-agenda/>
- Simon, P. (1964). *The sounds of silence*. Retrieved from <https://www.ultratop.be/nl/song/8947/Simon-&Garfunkel-The-Sounds-Of-Silence>
- United Nations. (2008). *United Nations declaration on the rights of Indigenous people*. Retrieved from https://www.un.org/development/desa/indigenouspeoples/wp-content/uploads/sites/19/2018/11/UNDRIP_E_web.pdf
- Waapalaneexkweew (N. R. Bowman-Farrell, Mohican/Lunaape). (2018). Looking backward but moving forward: Honoring the sacred and asserting the sovereign in Indigenous evaluation. *American Journal of Evaluation*, 39(4), 543–568. <https://doi.org/10.1177/1098214018790412>
- Waapalaneexkweew (N. Bowman, Mohican/Lunaape), & Dodge-Francis, C. (2018). Culturally responsive indigenous evaluation and tribal governments: Understanding the relationship. In F. Cram, K. A. Tibbetts, & J. LaFrance (Eds.), *Indigenous evaluation. New Directions for Evaluation*, 159, 17–31. <https://doi.org/10.1002/ev.20329>
- Wehipeihana, N., & McKegg, K. (2018). Values and culture in evaluative thinking: Insights from Aotearoa New Zealand. In A. T. Vo & T. Archibald (Eds.), *Evaluative thinking. New Directions for Evaluation*, 158, 93–107. <https://doi.org/10.1002/ev.20320>

AUTHOR INFORMATION

Nicole Bowman (Waapalaneexkweew) is a researcher and evaluator with the Wisconsin Center for Education Research at the University of Wisconsin – Madison and also owns a consulting firm, Bowman Performance Consulting, in Shawano, Wisconsin. She uses traditional Indigenous knowledge, systems, and culturally responsive evaluation methods to empower Tribal and non-Tribal agencies, communities, and people. Over the past two decades, she has worked with hundreds of programs nationally and internationally to increase their outcomes, capacities, and competencies, with the ultimate goal of more effective, responsive, and impactful “Nation-to-Nation” programming. She currently holds appointed positions for several professional and academic groups under the United Nations, American Evaluation Association, and the Center for Culturally Responsive Evaluation and Assessment, among others.

White Privilege and the Decolonization Work Needed in Evaluation to Support Indigenous Sovereignty and Self-Determination

Kate McKegg
The Kinnect Group

Abstract: *This paper builds on a keynote paper presented at the 2018 Canadian Evaluation Society annual conference by Kate McKegg, a Pākehā, non-Indigenous evaluator from Aotearoa, New Zealand. Kate reflects on the concept and implications for Indigenous people of white privilege in colonized Western nations. She discusses some of the ways in which white privilege and its consequences play out in the field of evaluation, perpetuating colonial sentiments and practices that maintain and reinforce inequities and injustice and potentially threaten the social justice aspirations of the field. Kate argues that those with white privilege have much work to do, unpacking and understanding their privilege if they are to have any chance of playing a role in deconstructing and dismantling the power structures that hold colonizing systems in place. She suggests that for evaluators to be effective allies for Indigenous sovereignty and self-determination, they must undertake ideological, cultural, emotional, and constitutional work. This work will be tough and scary and is not for the faint hearted. But it is vital to unlocking the potential transformation that can come from just and peaceful relationships that affirm and validate Indigenous peoples' ways of knowing and being.*

Keywords: *allies, colonization, evaluation, Indigenous, power, white privilege*

Résumé : *Le présent article vient la conférence invitée de Kate McKegg, une évaluatrice non autochtone pākehā d'Aotearoa en Nouvelle-Zélande, lors du Congrès annuel de la Société canadienne d'évaluation en 2018. Mme McKegg discute du concept et des conséquences, pour les personnes autochtones, du privilège blanc dans les Nations occidentales colonisées. Elle parle des façons par lesquelles le privilège blanc et ses conséquences jouent un rôle dans le domaine de l'évaluation, perpétuant ainsi des pratiques et des sentiments coloniaux qui maintiennent et renforcent des inégalités et des injustices, et menacent potentiellement les aspirations en matière de justice sociale de ce domaine. Mme McKegg fait valoir que les personnes bénéficiant du privilège blanc ont beaucoup de travail à faire pour décortiquer et comprendre leurs privilèges, si elles veulent avoir l'occasion de jouer un rôle dans la déconstruction et le démantèlement des structures de pouvoir qui permettent aux systèmes colonisateurs*

Corresponding author: Kate McKegg, 51a Bridle Creek Road, RD 1, Raglan, 3295, New Zealand; kmckegg@me.com

de se maintenir en place. Elle suggère que si les évaluateurs et les évaluatrices veulent être des alliés efficaces pour la souveraineté et l'autodétermination autochtones, ils et elles doivent y aller d'efforts idéologiques, culturels, émotionnels et constitutionnels. Il s'agit d'un travail difficile et déstabilisant qui testera leurs limites. Mais il s'agit aussi d'un travail vital pour permettre la transformation potentielle qui découle de relations justes et paisibles qui affirment et valident les connaissances et l'existence des peuples autochtones.

Mots clé : *alliés, colonisation, évaluation, autochtone, pouvoir, privilège blanc*

My name is Kate McKegg.

On my father's side I descend from Irish, Scottish, and English ancestors who arrived in New Zealand on ships in 1840 and 1860—with the families settling in Wanaka, Otaki, and the Manawatu region of New Zealand. On my mother's side I descend from Irish and Scottish ancestors who arrived on ships in the 1860s and 1880s. These early settlers began life on the west coast of the South Island of New Zealand, moving to Napier and then later to Titirangi near Auckland.

My ancestors were weavers, publicans, writers, business people, health professionals, and social activists—handing down strong values about the importance of community and family, of self-reliance, fairness, hard work, and education. They all came to New Zealand for a better life than the one they had—much like the millions of others who migrated from Europe in the nineteenth century. I identify as a Pākehā New Zealander. Pākehā in New Zealand roughly means those who descend from immigrant settlers from the United Kingdom and Europe. What we have in common with each other is the privilege of being beneficiaries of a colonization process in Aotearoa, New Zealand. Our identity as Pākehā settlers has been forged in relation to Māori (the Indigenous people of Aotearoa, New Zealand), developing over time, through the process of colonization (Bell, 2014; Hotere-Barnes, 2015).

I begin with this personal background and history because I believe it's important for those of us who are non-Indigenous, living in colonized lands, to recognize that in all our relationships with Indigenous people in colonized countries, there are power relations that are historical, political, and economic. These relations need to be carefully considered in all the work that we do as evaluators (Berghan et al., 2017). My introduction is intended to locate myself with respect to my ancestors and to the land, as a colonial settler. It models a process of introduction we use at home in New Zealand called *mihimihi* (a traditional Māori introduction).

As I prepared for the keynote panel presentation at the Canadian Evaluation Society Conference in Calgary in 2018, I pondered the following question asked many years ago by Ernie House: How do we ensure that the evaluation is “socially just as well as true, that it attends to the interests of everyone in society and not solely the privileged” (House & Howe, 1999, p. 244). I reflected specifically on the hegemonic forces of power and privilege at play for those of us who inhabit

non-Indigenous, white settler identities (Huygens, 2011) because these forces may well threaten the contribution that evaluation could make to a world of the future that, in the words of Paulo Freire (2000, p. 26), is “less ugly, more beautiful, less discriminatory, more democratic, less dehumanizing, and more humane.”

The purpose of my keynote presentation was to present a non-Indigenous perspective on what it might mean to be a non-Indigenous evaluator and an ally in promoting sovereignty and self-determination for Indigenous people.

REFLECTING ON WHITE PRIVILEGE

I have reflected on Donna Mertens’s comments from 2001 that “[d]espite the application of outstanding minds in the evaluation field” from the earliest days of evaluation as we know it, “the social problems that currently confront us globally, and thereby provide on-going justification for the need for evaluation, are quite similar to those that instigated the first calls for evaluation decades ago” (Mertens, 2001, p. 367). We seem to be faced with some of the most pressing and urgent “wicked” problems that humanity has ever faced, such as homelessness, inequality, workforce changes, biodiversity loss, water degradation, and climate change (McKegg, 2013). Along with these global issues, the concerns and voices of Indigenous and minority communities are growing louder about the failure of democratic ideals, such as participation, equity, and social justice (McKegg, 2013).

What role should evaluation and evaluators play in relation to these concerns and voices? For those of us who consider evaluation to be a democratizing practice—with the goal of social betterment—it is a critical time for us to consider our location, role, and stance in relation to the pursuit of key democratic ideals such as equity and social justice. Karen Kirkhart (2015, p. 11) remarked recently on a visit to New Zealand that the evaluation profession values equity and social justice, and that strong evaluation is “evaluation that honours and advances these values in the questions it raises, the evidence it gathers, the relationships it builds, and the privilege it carries.” She noted that evaluation and evaluators carry privilege, and that failing to acknowledge the dynamics of power and privilege undercuts evaluation’s ability to advance social justice.

The concept of privilege is complicated. There are all kinds of privilege, and they intersect. Often, it is unexamined and invisible to those of us that have it (Kirkhart, 2015). In this paper, I am talking about white privilege. All white people in Western colonized countries have white privilege. Not all are racist, but all benefit from the privilege of whiteness in a system that assumes whiteness is normal. Mostly, whiteness isn’t really talked about. White people in Western nations don’t even think of themselves as white; they are just people. Everyone else is an “other”: a Māori man, or a Black woman, or an Asian child. Whiteness is connected to economic power and class—and is probably least understood by those it privileges (Milne, 2009). A final clarification—when I refer to white settlers, I am referring to those of us who descend from the white settlers of colonized Western nations.

My interest in talking about white settler privilege stems from frustration—frustration with the inertia of a dominant socio-cultural, educational, and economic system set up to privilege white settler colonizers (Huygens, 2011) to shift and address deep-rooted, intergenerational injustices that continue to be inflicted on Indigenous people in my own country and other wealthy, colonized nations. In some of the wealthiest nations in the world, such as Canada, New Zealand, Australia, and the United States, Indigenous people are some of the poorest citizens (Cornell, 2006). Cornell (2006, p. 1) argues that “the wealth of these countries has been built substantially on resources taken from these peoples, whose poverty—in the grand scheme of things—is a recent creation.”

EVALUATION AND WHITE PRIVILEGE

The field of evaluation is not immune to having a role in maintaining the dominant system of white privilege and power in Western nations. Evaluation theory and practice is dominated by Western evaluation frameworks that continue to systematically exclude and fail to recognize culturally distinct ontologies and epistemologies (Masters-Awatere & Nikora, 2017). I see every day the perpetuation of relations of power and white privilege in evaluation. I would argue that attempts and good intentions by the evaluation field to address social justice and equity concerns have largely failed to respond appropriately to these calls for our field to adapt and transform to address the very real needs of Indigenous people for equity, social justice, and sovereignty. The voices of our indigenous colleagues are the evidence I use for this stance (Wehipeihana, 2008; Waapalaneexkweew [Bowman-Farrell], 2018).

Even evaluation approaches that are called participatory, collaborative, and/or culturally responsive have for too long “been about providing space for non-Indigenous to do evaluation—some more than others. They have in the main been about providing the tips, passwords, and passport for non-Indigenous to do or facilitate evaluation in Indigenous communities and about processes that provide the trappings of permission for non-Indigenous evaluators to work in Indigenous spaces” (Wehipeihana, 2008). Participatory and collaborative evaluation processes, no matter how sensitively they are constructed—when they are controlled and done by white settlers—will not transcend the socio-cultural politics of our historical identities. They do quite the opposite: they perpetuate colonial sentiments (Bishop, 2011). One white New Zealand evaluator and researcher describes believing that these processes will be valid for Indigenous people as “stupid optimism” (Hotere-Barnes, 2015).

I believe that what our field’s attempts have consistently failed to recognize and acknowledge, however well intentioned they are, is that they are largely conceived and controlled by non-Indigenous commissioners and evaluators—with Indigenous involvement ranging from token through to infrequent, with occasional genuine partnership. With the power and control of evaluation in the hands of the dominant cultural paradigm, we simply will not be transformative.

So even when (and perhaps because) there are good intentions to do the right thing, there are no fundamental shifts in power imbalances between Indigenous and non-Indigenous peoples or the systems within which we operate. Unwittingly, our attempts to do the right thing may actually have served to further displace, overshadow, and mask Indigenous evaluation efforts.

It is encouraging to me that we live at a time where Indigenous populations across the world are exploring ways to decolonize, indigenize, and re-imagine knowledge theory and practice in every academic discipline and practice that is informed by their worldviews—including evaluation (Chilisa, 2012). Many colonized and marginalized peoples are working to claim back lost identities, and as evaluator and author Bagele Chilisa says, they are working to create spaces for significant selfhoods as well as writing back and talking back to the West in modes couched in their own histories, cultures, and linguistic and life experiences (Chilisa, 2012).

As privileged white settler evaluators, we are thus presented with an opportunity to challenge ourselves to unpack the invisible knapsack of white privilege (McIntosh, 1988) that often stifles our potential to affirm and validate the knowledge, philosophies, and worldviews of the diverse communities in which we live and work. In the unpacking, we will have to do more than simply point to the ideas; we will have to learn to *inhabit* them (Meyer, 2011). It is only by inhabiting the deconstruction of white privilege—in a truly holistic way—that we stand a chance of navigating the tricky ground ahead of us (Smith, 2005), of revealing, acknowledging, and dismantling the hidden, unseen, and colossal power that is held firmly in place by those with white privilege.

UNPACKING WHITE PRIVILEGE

Over some years, my own journey of unpacking my evaluation practice has evolved to encompass an overtly political, values-based orientation that includes affirming Māori sovereignty, as it is set out in *Te Tiriti o Waitangi*,¹ as well as an ethical imperative to put Māori aspirations at the heart of any evaluation that includes, is about, is for, or is done with or as Māori (Wehipeihana, current volume). I have come to recognize that as a Pākehā settler, I am a colonizer and that we still live in a colonizing society. Colonization is not consensual, nor is it over (Margaret, 2018). Institutional racism and culturally unsafe practices are still the normal way to do things (Berghan et al., 2017, p. 20).

My white privilege (and that of all other Pākehā white settlers) comes from the myriad ways in which my Pākehā white settler ancestors benefited from the systems they put in place for their ways—such as language, education, legal systems, and other norms—at the same time denying Māori the basic right to live and express themselves on their lands, in *their* ways (Margaret, 2018). My privilege is current both because of intergenerational transmission and because colonization is an ongoing activity. The structures of colonization have not been dismantled.

If there is to be social justice for Indigenous people and communities, in evaluation as well as more generally, the imperative is for those of us who are white, privileged settlers to change.

BECOMING A USEFUL ALLY

Becoming a useful ally for Indigenous sovereignty and social justice is not a “self-appointed identity” or a badge we acquire, and it’s not something we should take on out of self-interest or ego. It is a role that those committed to a collective struggle for social justice take on, for life (Swiftwolfe & Shaw, n.d.). The work involved is a journey full of complexity and uncertainty. It can be exhilarating and also deeply challenging because our dominant knowledge systems and structures are powerful and resistant to allowing the necessary time and space for equitable co-existence of other knowledges (Wehipeihana & McKegg, 2018).

The journey of becoming an ally for Indigenous social justice in any context, including evaluation, requires that we shift our gaze from helping others, determining needs for others, designing evaluations for others, to changing ourselves (Huygens, 2011; McKegg, 2013; Smith, 2012). Only when we understand our histories, and our collusion as perpetrators of colonization, and the privilege this affords us, can we then move on to the journey of becoming useful allies. So, as well as supporting the struggles of Indigenous peoples, we must first engage in separate and specific work amongst ourselves (Margaret, 2013, p. 6). The reality is that most of us who identify as white settlers don’t know our own histories, let alone those of Indigenous people, nor our place in them. Margaret (2018, p. 3) argues that “we don’t know ourselves, our white ways, and we don’t know the ways of [the] land. We are poorly equipped to act well in the relationships that allow us to be [on the lands we occupy].”

It’s hard to hear the real history; it’s not like the one we’ve been telling ourselves to elevate and justify our positions of privilege. We didn’t discover the lands we occupy; the people who were on the lands our ancestors settled didn’t benefit from our occupation. We benefited, and continue to do so, at the expense of Indigenous people’s cultures, language, economies, and spiritualities.

The following quotation by a New Zealand treaty advocate and educator sums up the practice of working as an ally: “The practice of working as an ally is contextual and relational therefore issues need to be worked through with regard to the specific circumstances and relationship/s. There is not a simple checklist that allows you to ‘be a good ally’” (Margaret, 2010, p. 6).

We (those of us who are white settlers and colonizers) must embrace a journey of reconciliation with ourselves. We have to find a way to reconcile our collective cultural cognitive dissonance—between what we’ve told ourselves and believed wholeheartedly was “the story” of settlement, and what we now know has been the impact of our settlement on Indigenous people. We have to begin with knowing and accepting ourselves, our histories, our privilege and power. Only by

knowing ourselves can we begin unravelling the intersections between privilege, power, colonization, and racism.

Our readiness to do this work as evaluators is emergent, and we are just beginning the conversations. For Indigenous communities, their experience of evaluation to date has been highly problematic (Blanchet-Cohen, Geoffroy, & Hoyos, 2018). The use of externally imposed frameworks, criteria, and methods, and the resulting judgments, have had little meaning or value for Indigenous communities; worse, they have perpetuated harmful systems and structures. How we cultivate evaluation practice that is able to respectfully and humbly acknowledge, value, and honour diverse knowledge traditions and paradigms is still, for the most part, an aspiration yet to be realized (Wehipeihana & McKegg, 2018).

DECOLONIZATION WORK

If evaluation is, at the very least, to stop doing harm to Indigenous people, we have to begin our own work of conscious decolonization (Berghan et al., 2017; Blanchet-Cohen et al., 2018; Waapalaneekweew [Bowman-Farrell], 2018). This work is not for the faint-hearted; it involves intentional ideological, cultural, emotional, and constitutional work (Huygens, 2011).

Ideological work involves critically revisiting *and then retelling* the history of our relationship with Indigenous peoples in order to appreciate that Western, colonial ideologies have shaped all our worldviews and have been self-legitimizing for those of us with white privilege (Huygens, 2011). This ideological work points white settler evaluators toward the need to recognize and acknowledge Indigenous knowledge systems and ways of knowing. And it challenges white settler evaluators to consider the adequacy of our Western theories, frameworks, and knowledge for all evaluation work done for, with, or by Indigenous people (McKegg, 2013; Wehipeihana & McKegg, 2018).

Cultural work involves critiquing those aspects of our identity, culture, and tradition that will *not* serve the sovereignty and self-determination of Indigenous people. This implies that we need to rethink and reconsider the roles we play on evaluation projects and teams. Privileging Indigenous voices, values, and traditions is a stance that those of us who are white settlers must be prepared to take if we are to shift the balance of power (McKegg, 2013). This work also points to the kind of relationships we need to learn how to form in our evaluation work, allowing Indigenous people to lead, with us following—where we make time for trust to develop so that we can identify racism and oppressive practice and strategize with each other to take action (Berghan et al., 2017). This is also particularly important work in those spaces that are not considered Indigenous. We can and we must learn to challenge the perpetuation of racism, wherever and whenever we see it.

This is also emotional work. It involves standing up and challenging white privilege, even when it's scary (Swiftwolfe & Shaw, n.d.). As we bring to the notice of those with white privilege our collective complicity and ignorance, we can expect to have to deal with feelings of shock, fear, guilt, denial, and so on. And when

this happens, we should not reach for assistance from those we have oppressed. When we ask Indigenous folks to take the time to sit down and educate us on the specific issues that they face and how we can be better allies, rather than doing the work of educating ourselves, we are making it about us. In evaluation situations, our job is to challenge others like us about their power over an evaluation process, design, system, or practice when it has implications for perpetuating inequalities and oppression (McKegg, 2013).

Constitutional work involves the privileged being responsive in practice and form to the aspirations, perspectives, and views of Indigenous people. This means bringing about changes to our policies, procedures, structures, and systems, as well as resource flows. It raises questions for us about how decisions are made throughout evaluation processes. For example, how are the needs and resources for Indigenous evaluation determined? How might we commission and contract for evaluation in ways that shift and disrupt the power and dominant forces that maintain our current practice?

Choosing to embark on this journey will be unsettling, and it is likely that feelings of shame, discomfort, and uncertainty will part of the experience. Working the spaces in between Indigenous and white settlers requires joint commitment to creating conditions where disagreements, emotional flux, and emergence are to be expected and normal (Cram & Phillips, 2012). Arriving at this recognition takes intellectual, emotional, cultural, and constitutional work over time, which will be challenging (Huygens, 2011). Yet, as Hotere-Barnes (2015) says, to ignore these difficulties perpetuates and preserves racist systems. If we work to shift our thinking and understandings of ourselves and our relationships with Indigenous people, our subsequent actions and interactions are less likely to reinforce negative and unproductive patterns and outcomes (Bishop, 2017). Learning to remain quiet, listen well, pay respectful attention, share power, and sustain relationships over the long term will also unlock powerful learning opportunities and fill your whole self, heart, head, soul, and spirit.

It will also support the unlocking of the innovation potential and the transformation we so desperately need in our thinking if we to reduce the world's suffering that we are all experiencing. The human and environmental suffering that is, in major part, the consequence of colonization surrounds us in rising mental illness and distress; homelessness; workforce, justice, health, and educational inequities; environmental degradation; and more. A catalyst for a sustainable future that is also more just, more equitable, more kind, more holistic can be found if we are prepared to shift our gaze and practice (Hudson, Roberts, Smith, Tiakiwai, & Hemi, 2012). Jen Margaret (2013, pp. 7–8) so aptly says that “[j]ust and peaceful relationships are critical to the health of communities, and knowledge of how to work as allies has the potential to contribute to positive relationships throughout society.”

I have a lifetime commitment to ongoing decolonization work in my personal and professional life in the hope that I can make a contribution to supporting Indigenous people's claims to sovereignty over their lives, including what kind of

evaluation takes place in all matters that directly or indirectly affect or concern them. I remain hopeful that there will come a time when Indigenous practices, languages, ways of being and knowing, rights, perspectives, and thinking are affirmed and realized, and when this happens, I'm confident it will benefit us all.

NOTE

- 1 *Te Tiriti o Waitangi* is a treaty signed in 1840 by over 500 Māori chiefs and representatives of the British monarch. It outlined the terms and conditions of settlement for settlers and reaffirmed Māori sovereignty. It is the closest document New Zealand has to a written constitution (Berghan et al., 2017).

REFERENCES

- Bell, A. (2014). *Relating Māori and Pākehā: The politics of Indigenous and settler identities*. Palmerston North, New Zealand: Massey University.
- Berghan, G., Came, H., Coupe, N., Doole, C., Fay, J., McCreanor, T., & Simpson, T. (2017). *Tiriti-based health promotion practice*. Auckland, New Zealand: STIR.
- Bishop, R. (2011). *Freeing ourselves*. Transgressions: Cultural Studies and Education, 66. Rotterdam, Netherlands: Sense Publishers.
- Bishop, R. (2017). Relationships are fundamental to learning. *Principal Connections*, 20(3). Retrieved from <http://cognitioneducation.com/wp-content/uploads/2017/08/Russell-Bishop-Editorial-for-CPCO.pdf>
- Blanchet-Cohen, N., Geoffroy, P., & Hoyos, L. M. (2018, October). Seeking culturally safe developmental evaluation: Supporting the shift in services for indigenous children. *Journal of MultiDisciplinary Evaluation*, 14(31), 19–31.
- Chilisa, B. (2012). *Indigenous research methodologies*. Los Angeles, CA: SAGE.
- Cornell, S. (2006). *Indigenous peoples, poverty and self-determination in Australia, New Zealand, Canada and the United States*. JOPNA No. 2006-02. Retrieved from <https://hpaied.org/sites/default/files/publications/Indigenous%20Peoples,%20Poverty%20and%20Self%20Determination%20in%20Australia,%20New%20Zealand,%20Canada%20and%20the%20US.pdf>
- Cram, F., & Phillips, H. (2012). Claiming interstitial space for multicultural, transdisciplinary research through community-up values. *International Journal of Critical Indigenous Studies*, 5(2), 36–49. <https://doi.org/10.5204/ijcis.v5i2.89>
- Freire, P. (2000). *Pedagogy of the oppressed*. New York, NY: Herder & Herder.
- Hotere-Barnes, A. (2015). Generating “non-stupid optimism”: Addressing Pākehā paralysis in Māori educational research. *New Zealand Journal of Educational Studies*, 50(1), 39–53. <https://doi.org/10.1007/s40841-015-0007-y>
- House, E. R., & Howe, K. (1999). *Values in evaluation and social research*. Thousand Oaks, CA: SAGE.
- Hudson, M., Roberts, M., Smith, L., Tiakiwai, S.-J., & Hemi, M. (2012). The art of dialogue with Indigenous communities in the new biotechnology world. *New Genetics and Society*, 31(1), 11–24. <https://doi.org/10.1080/14636778.2011.597979>

- Huygens, I. (2011). Developing a decolonisation practice for settler colonisers: A case study from Aotearoa New Zealand. *Settler Colonial Studies*, 1(2), 53–81. <https://doi.org/10.1080/2201473x.2011.10648812>
- Kirkhart, K. E. (2015). Unpacking the evaluator's toolbox: Observations on evaluation, privilege, equity and justice. *Evaluation Matters—He Hake Tō Te Aromatawai*, 1, 7–24. <https://doi.org/10.18296/em.0002>
- Margaret, J. (2010). *Working as allies: Winston Churchill fellowship report*. Retrieved from <http://www.awea.org.nz/sites/awea.org.nz/files/Jen%20Margaret%20Winston%20Churchill%20Report%202010.pdf>
- Margaret, J. (2013). *Working as allies: Supporters of indigenous justice reflect*. Auckland, New Zealand: Auckland Workers Educational Association.
- Margaret, J. (2018). *Ka pū te ruha, ka hao te rangatahi: Change in the Pākehā Nation*. State of the Pākehā Nation. Retrieved from <https://groundwork.org.nz/resources/>
- Masters-Awatere, B., & Nikora, L. W. (2017). Indigenous programmes and evaluation: An excluded worldview. *Evaluation Matters—He Take Tō Te Aromatawai*, 3, 1–27. <https://doi.org/10.18296/em.0020>
- McIntosh, P. (1988). *White privilege and male privilege: A personal account of coming to see correspondences through work in women's studies*. Working Paper No. 189. Wellesley, MA: Wellesley Centers for Women.
- McKegg, K. (2013). Evaluation responsibility and leadership in the face of failing democracies. *American Journal of Evaluation*, 34(4), 579–582. <https://doi.org/10.1177/1098214013494458>
- Mertens, D. M. (2001). Inclusivity and transformation: Evaluation in 2010. *American Journal of Evaluation*, 22(3), 367–374. [https://doi.org/10.1016/S1098-2140\(01\)00150-3](https://doi.org/10.1016/S1098-2140(01)00150-3)
- Meyer, M. (2011). *Holographic epistemology: Native common sense*. Unpublished Paper.
- Milne, A. (2009). *Colouring in the white spaces: Cultural identity and learning in school*. Auckland, New Zealand: APPA.
- Smith, L. T. (2005). On tricky ground, researching the native in the age of uncertainty. In N. Denzin & Y. S. Lincoln (Eds.), *The SAGE handbook of qualitative research* (pp. 1–12). Thousand Oaks, CA: SAGE.
- Smith, L. T. (2012). *Decolonising methodologies* (2nd ed.). London, England: Zed Books.
- Swiftwolfe, D., & Shaw, L. (n.d.). *Indigenous ally toolkit*. Montreal Indigenous Community Network. Retrieved from http://reseamtlnetwork.com/wp-content/uploads/2018/12/Ally_web-2.pdf?fbclid=IwAR2Wbpz_SJ78lZg24nJEB0kY3I1s5sCaIx6GnJjwedvcB-7vDjnOwie-wrMI
- Waapalaneeckweew [N. Bowman-Farrell]. (2018). Looking backward but moving forward: Honoring the sacred and asserting the sovereign in Indigenous evaluation. *American Journal of Evaluation*, 39(4), 543–568. <https://doi.org/10.1177/1098214018790412>
- Wehipeihana, N. (2008). Indigenous evaluation: A journey beyond participatory and collaborative approaches in evaluation. Presentation at the Australasian Evaluation Society Conference, Perth.
- Wehipeihana, N., & McKegg, K. (2018). Values and culture in evaluative thinking: Insights from Aotearoa New Zealand. *New Directions for Evaluation*, 18(158), 93–106. <https://doi.org/10.1002/ev.20320>

AUTHOR INFORMATION

Kate McKegg is an independent evaluation consultant with over 25 years' evaluation experience. She is the director of The Knowledge Institute Ltd (www.knowledgeinstitute.co.nz), co-founder of the Developmental Evaluation Institute (<https://developmental-evaluation.org>), and a member of the Kinnect Group (www.kinnect.co.nz) and the Tuakana Teina Collective. Kate is a founding member and past convenor of the Aotearoa New Zealand Evaluation Association (www.anzea.org.nz). She is co-editor of New Zealand's only evaluation text, *Evaluating Policy and Practice: A New Zealand Reader* (2003), and co-editor (along with Michael Quinn Patton and Nan Wehipeihana) of the book *Developmental Evaluation: Real World Applications, Adapted Tools, Questions Answered, Emergent Issues, Lessons Learned, and Essential Principles* (Guilford Press, 2015).

Increasing Cultural Competence in Support of Indigenous-Led Evaluation: A Necessary Step toward Indigenous-Led Evaluation

Nan Wehipeihana

Research Evaluation Consultancy Limited

Abstract: *This paper builds on a keynote paper presented at the 2018 Canadian Evaluation Society annual conference by Nan Wehipeihana, an Indigenous (Māori) evaluator from Aotearoa New Zealand. Nan defines Indigenous evaluation as evaluation that is led by Indigenous peoples; has clear benefits for Indigenous peoples; has Indigenous people comprising most of the evaluation team; is responsive to tribal and community contexts; and is guided and underpinned by Indigenous principles, practices, and knowledge. She argues for Indigenous led as a key criterion for Indigenous evaluation, with no assumed or automatic role for non-Indigenous peoples unless by invitation. She outlines a range of tactics to support the development of Indigenous evaluators and Indigenous evaluation and presents a model for non-Indigenous evaluators to assess their practice and explore how power is shared or not shared in evaluation with Indigenous peoples, as a necessary precursor to increasing control of evaluation by Indigenous peoples.*

Keywords: *control, culturally responsive evaluation, decision making, Indigenous evaluation, Indigenous-led, Kaupapa Māori, power*

Résumé : *Cet article vient compléter la conférence invitée de Nan Wehipeihana, une évaluatrice autochtone (māori) d'Aotearoa en Nouvelle-Zélande, lors du Congrès annuel de la Société canadienne d'évaluation en 2018. Mme Wehipeihana définit l'évaluation autochtone comme une évaluation menée par des personnes autochtones ; qui présente des avantages clairs pour les peuples autochtones; dont l'équipe d'évaluation est surtout composée de personnes autochtones ; qui est consciente des contextes communautaires et tribaux ; et qui est orientée par des principes, des pratiques et des connaissances autochtones. Elle soutient que l'évaluation autochtone devrait être menée par des Autochtones, sans rôles présumés ou automatiques pour les personnes non autochtones à moins qu'elles ne soient invitées à y participer. Elle décrit toute une gamme de tactiques visant à appuyer le développement des évaluateurs et des évaluatrices autochtones et de l'évaluation autochtone. Elle présente aussi un modèle permettant aux évaluateurs et aux évaluatrices non autochtones d'évaluer leur pratique et d'explorer la manière dont le pouvoir est partagé ou non, en évaluation,*

Corresponding author: Nan Wehipeihana, P.O. Box 6545, Wellington 6141, New Zealand;
nan.wehipeihana@gmail.com,

avec les peuples autochtones, en tant que préalable nécessaire à l'augmentation du contrôle exercé sur l'évaluation par les peuples autochtones.

Mots clé : *contrôle, évaluation adaptée à la culture, prise de décision, évaluation autochtone, leadership autochtone, Kaupapa Māori, pouvoir*

<i>Ko Taraura te pae maunga.</i>	<i>Tararua are the mountain ranges</i>
<i>Ko Ohau tōku awa.</i>	<i>Ohau is my river</i>
<i>Ko Tukorehe tōku marae.</i>	<i>Tukorehe is my traditional meeting place</i>
<i>Ko Ngāti Tukorehe te iwi i te taha o tōku matua.</i>	<i>Ngāti Tukorehe is my tribe on my father's side.</i>
<i>Ko Ngāti Porou me Te-Whānau-ā-Apanui ōku iwi i te taha ūkaipō.</i>	<i>Ngāti Porou and Te-Whānau-ā-Apanui are my tribes on my mother's side.</i>
<i>Tuku mihi ōku maunga ki ngā maunga o tēnei rohe, o tēnei whenua.</i>	<i>My mountains greet the mountains of this region, of this land</i>
<i>Tuku mihi ōku awa, ōku moana ki ngā awa me ngā moana o tēnei rohe, o tēnei whenua.</i>	<i>My rivers and my oceans greet the rivers and oceans of this region, of this land.</i>
<i>Ki te tangata whenua o tēnei whenua, kei te mihi.</i>	<i>I acknowledge the traditional people of this land.</i>
<i>Ki te atua i runga rawa, nāna i homai nāna i tango, kei te mihi.</i>	<i>To the creator of all things, who gives and takes all things, I greet you.</i>
<i>I ngā mate kua rūpeke atu ki te pō, haere haere oki oki atu.</i>	<i>To the people who have gone before us, I acknowledge you.</i>
<i>Ko Ngahorihori (Nan) Wehipeihana ahau.</i>	<i>I am Ngahorihori (Nan) Wehipeihana</i>
<i>Tuku mihi ki a koutou katoa, no ngā hau e whā, tēnā koutou katoa.</i>	<i>And to all gathered here today, from the four winds, I greet you.</i>

I start my presentation today with a *mihimihi* (a traditional introduction), as I start every presentation, whether at home in Aotearoa New Zealand or as a visitor on another people's land. I do this by identifying my connections to places of significance to my people, our mountains, rivers, and *marae* (traditional meeting places) and I acknowledge my relationships to people both past and present. I also greet the traditional custodians of this land, attesting to my visitor status.

My *mihimihi* affirms my *whakapapa* (genealogy and identity) and my links to the land, to the natural environment, and to ancestors, the spirits, and the cosmos (Marsden & Royal, 2003). It reminds us of our connectedness to people and to place (Henare, 1988; LaFrance & Nichols, 2010), that we all connect to a spiritual source and exist in a web of reciprocity with each other and all creation (Spiller & Stockdale, 2013).

Entering Māori communities, like many Indigenous communities, typically involves other rituals of encounter such as *pōwhiri* (formal welcome), *whaikōreoro*

(speeches), and *karakia* (ritual chants or prayer). Collectively, these cultural protocols signal a movement from the contemporary world dominated mainly by Western practices into a Māori world where Māori values and norms hold sway; and they help to prepare and settle us for the context in which we will be working (Wehipeihana & McKegg, 2018).

In this presentation, I define Indigenous evaluation, outline the rationale for Indigenous-led, put forward a strategy to support Indigenous-led evaluation, and share a model (Wehipeihana, 2013) for non-Indigenous evaluators to reflect on their positioning and ways of working to reveal the power dynamics that are a barrier to Indigenous evaluation being led by Indigenous peoples.

DEFINING INDIGENOUS EVALUATION

When I think about Indigenous evaluation (or in my case Kaupapa Māori evaluation), it is about evaluation by Indigenous, for Indigenous, with Indigenous, and as Indigenous; and where there is no assumed role for non-Indigenous people, unless by invitation. In Aotearoa New Zealand, “By Māori, for Māori, with Māori” (Cram, 2016; Cram, Chilisa, & Mertens, 2013) has variously been part of the research and evaluation agenda since the 1980s (and more recently “as Māori”: Durie, 2001; Wehipeihana, McKegg, Thompson, & Pipi, 2015) (see Table 1).

INDIGENOUS SELF-DETERMINATION IS FUNDAMENTAL

Kaupapa Māori evaluation has its foundations in the principle of *Tino Rangatiratanga* (chiefly leadership) which appears in the Māori version of the Treaty of Waitangi

Table 1. Core determinants of Māori and Indigenous evaluation

By Māori: (by Indigenous)	is about evaluation led by Māori and where Māori have the overall authority and power to make decisions about the evaluation design, methods, evaluative criteria, and ways of working
For Māori: (for Indigenous)	is about ensuring there are clear benefits for Māori from the evaluation, and Māori aspirations are acknowledged within the evaluation
With Māori: (with Indigenous)	is where Māori make up the majority of the evaluation team; the evaluation is responsive to tribal and community contexts and respectfully observes and utilizes <i>te reo Māori me ngā tikanga</i> (Māori language and cultural practices)
As Māori: (as Indigenous)	is about evaluation that is guided, informed, and underpinned by <i>kaupapa tuku iho</i> (cultural values gifted by ancestors) and <i>tikanga Māori</i> (Māori cultural practices)
Role of non-Māori (non-Indigenous)	is where there is no automatic or assumed role for non-Māori on the evaluation team and the participation of non-Māori people is by invitation only

signed by the British Crown and *rangatira* (chiefs) in 1840. *Tino Rangatiratanga* has come to be understood as a translation for the term “self-determination” and refers to determination by Māori of issues that have an impact on Māori. *Tino Rangatiratanga* has been at the forefront of Māori aspirations since the signing of the treaty in 1840 and remains so today.

Kaupapa Māori evaluation also draws on Kaupapa Māori theory (Smith, 1997) and its use as a strategic and political tool to advance Māori aspirations and agendas. In Aotearoa, Kaupapa Māori theory is the touchstone for research and evaluation with Māori. Cram (2001) defines Kaupapa Māori as a Māori way, underpinned by Māori philosophies, values, and principles that have an aspirational orientation or vision for Māori cultural, social, and economic wellbeing.

At the same time, “By Māori, for Māori” emerged as part of the political and aspirational discourse of self-determination around Treaty of Waitangi claims and Kaupapa Māori theory. “By Māori, for Māori” was a cultural and political position that asserted the right of Māori to lead their own development, that is, be self-determining. Māori expectations in relation to the conduct of research and evaluation began to be reflected in Māori-specific ethical guidelines, such as those developed by the Health Research Council of New Zealand (2010) and by government agencies such as Te Puni Kōkiri (1999)—the Ministry for Māori Development—and the Ministry for Social Development (2004). However, the role and appropriateness of Pākehā¹ researchers in research involving Māori was keenly debated. On the one hand, you had Māori arguing there was no place for Pākehā in Māori research (Walker, 1990). On the other hand Smith (1990) identified four culturally appropriate research models to address Māori research needs. These models² of research were seen to have merit at the time, to mitigate the lack of Māori research capacity and capability (Cram, 1997).

Conducting evaluation within a Kaupapa Māori framework means that non-Māori evaluators undertake evaluation at the invitation of, and in partnership with, the Māori community (Mertens, 2009). For non-Māori evaluators, their participation in Kaupapa Māori evaluation is akin to the notion of guesthood (Harvey, 2003). From the outset, being a guest signifies acceptance of being on someone else’s “turf” and a willingness to defer to the protocols of one’s host. Within this context, guests understand that their role will be determined by Māori, and, ideally, they will have a heightened sense of awareness of their positionality and of being an “outsider” in the evaluation (Brayboy & Deyhle, 2000).

So my definition of Indigenous evaluation has its roots in the struggle by Māori to be self-determining and to exercise control and decision making over their lives and the things that matter to Māori. Māori, however, are not alone in asserting a self-determination agenda. Indigenous people across the globe have long advocated for self-determination, and we see the importance of Indigenous self-determination reflected in Articles 3 and 4 of the United Nations Declaration on the Rights of Indigenous Peoples Resolution (adopted by the General Assembly on September 13, 2007) (see Table 2).

Table 2. United Nations Declaration on the Rights of Indigenous Peoples: Articles 3 and 4

Article 3: Indigenous peoples have the right to self-determination. By virtue of that right they freely determine their political status and freely pursue their economic, social and cultural development.

Article 4: Indigenous peoples, in exercising their right to self-determination, have the right to autonomy or self-government in matters relating to their internal and local affairs, as well as ways and means for financing their autonomous functions.

WHY INDIGENOUS-LED EVALUATION?

Much has been written about the location of research and evaluation within Western traditions, including the dominance of this scholarship (Bishop, 2005; Smith, 1999; Wehipeihana & McKegg, 2018); the role of imperialism, colonialism, and globalization in the elevation and privileging of Western research traditions (Chilisa, 2012; Cram, 2009); and research and evaluation as an exercise in power and political contestation for knowledge and resources (Bishop, 2005; Teariki, Spoonley, & Tomoana, 1992; Te Awekotuku, 1991).

If evaluation is to make a positive difference in the lives of Indigenous peoples, and to the things that matter for Indigenous peoples, then the dominant Western discourse needs to be held at bay. Indigenous-led, Indigenous control is essential if Indigenous values, principles, and ways of working are to hold sway in Indigenous evaluation and for the creation of a safe space for Indigenous communities and evaluators to be Indigenous.

When an evaluation is led by Indigenous people, they are more likely to have the cultural capital for the following:

- to facilitate respectful engagement and the observance of cultural protocols; they know what matters to get relationships and the evaluation off to a good start (Durie, 2001);
- to utilize Indigenous knowledge, methods, and ways of working in an evaluation; they can provide a cultural “reading” or assessment of the cultural fit or appropriateness of data-collection methods and tools for Indigenous peoples (Goodwin, Sauni, & Were, 2015);
- to facilitate understanding of what value and goodness look like through an Indigenous worldview; they can take the lead in sense making and analysis to ensure that the richness, subtlety, and nuance of meaning are not lost in translation and ensure the cultural validity of the evaluative conclusions drawn (Kirkhart, 2010).

When I listen to Indigenous speakers around the world, and I read the scant but growing indigenous evaluation literature, the “by” Indigenous or Indigenous-led

criterion is not prominent, and Indigenous evaluation is conflated with culturally responsive methodologies. What this does, in my view, is to tacitly provide permission and/or by default suggest that it is “okay” for non-Indigenous evaluators to lead evaluation with Indigenous peoples.

In Aotearoa New Zealand, Māori aspirations for control over their own lives and for self-determination means that it is *not* okay for non-Māori to lead evaluation with *whānau* (extended family), *hapū* (clan or sub-tribe), *iwi* (tribe), or Māori organizations. However, this is not the norm outside of Aotearoa New Zealand, where non-Indigenous evaluators mostly lead evaluation projects with and in Indigenous communities. To a large extent, this stems from unquestioned assumptions by the international evaluation community that they can undertake high-quality evaluations when working in Indigenous communities, or cross-culturally, where the culture, context, and language(s) are different from their own; collect good data; undertake culturally valid analysis; and make sound evaluative judgements (Wehipeihana, Davidson, McKegg, & Shanker, 2010). There are increasing challenges to this assumption, but Indigenous evaluator numbers are small, with limited influence, and decision-making power is typically vested in non-Indigenous funders and evaluators.

My definition of Indigenous evaluation has Indigenous-led as a critical criterion, with no assumed or automatic role for non-Indigenous peoples, unless by invitation. It reflects my personal and professional belief about what it takes to conduct high-quality evaluation with Indigenous peoples. While many things can be learned in relation to cultural knowledge, some things can be known and revealed only from within the culture. For these reasons, “there is no substitute for the cultural capital (understanding, knowledge, and intuit) that comes from being of the culture” (Wehipeihana et al., 2010, p. 188).

SUPPORTING INDIGENOUS-LED EVALUATION

One of the arguments I hear most often is that there aren’t enough Indigenous evaluators for evaluations with Indigenous peoples to be Indigenous-led. And I agree, so very simply more Indigenous evaluators are needed, and this needs to be a priority. Drawing on the Aotearoa New Zealand evaluation experience, some considerations for the Canadian Evaluation Society to support Indigenous-led evaluation include the following:

- supporting the development of Indigenous evaluators and Indigenous evaluation
- partnering with Indigenous people to guide evaluation
- increasing the cultural competence of non-Indigenous evaluators
- elevating Indigenous values and principles in the practice of evaluation in Canada
- understanding the importance of relationships when engaging with Indigenous peoples

- advocating for the credibility and inclusion of Indigenous values, methods, and criteria.

Supporting the Development of Indigenous Evaluators and Indigenous Evaluation

Community-based evaluation workshops, some explicitly targeting Māori, and funded through health-promotion and Māori development agencies have provided alternative pathways for Māori to enter the field of evaluation and for Māori to guide evaluation in their communities. Evaluation mentors, both Māori and non-Māori, have encouraged and supported Māori to complete tertiary evaluation qualifications.

The Aotearoa New Zealand Evaluation Association (ANZEA) has supported Māori (and Pacific) evaluators by providing and resourcing a dedicated convening space at its annual conference. Māori evaluators determine how this time is used and share how they have applied *matauranga Māori* (Māori knowledge systems), *te ao Māori* (Māori perspectives and worldviews), and tribal knowledge and cultural protocols in their evaluation practice. This contributes to the building of Kaupapa Māori evaluation practice and principles and the strengthening of a Māori evaluation network.

In the development of the Aotearoa New Zealand evaluator competencies and evaluation standards, there was rich conversation and dialogue about what matters to evaluation in New Zealand, and what quality and goodness means. Māori evaluators decided that one way of expressing this was to do it on their own terms; this led to the formation of Ma te Rae (the Māori evaluation association)—the first Indigenous evaluation organization in the world. Ma te Rae actively participates in EVAL Indigenous, has its own programme of development work, and sits comfortably alongside ANZEA.

From a base of less than five in 1999, Māori evaluators numbered approximately 50 in 2018.³

Partnering with Indigenous People to Guide Evaluation

Deep cultural knowledge is needed for evaluators to work safely and respectfully with Indigenous communities. This is not a skate-over-the-surface knowledge to get you in the door by observing a few cultural protocols. It is cultural knowledge that will help you respectfully navigate entry, agree on ways of working, including methods and approaches, and support Indigenous people to give expression to their views, values, and experiences.

Non-indigenous evaluators need to find an Indigenous community advisor or guide, such as a tribal Elder or community leader, to help guide evaluators and the evaluation. These cultural advisors provide historical and contextual information, connect evaluators to people in the community, and guide their observance of cultural protocols.

A word of caution: Even when you are Māori or Indigenous, you can't assume that your knowledge is applicable across all tribes and communities.

Increasing the Cultural Competence of Non-Indigenous Evaluators

First, do your homework before engaging with or entering Indigenous communities. Find out about the history and current context (Ormond, Cram, & Carter, 2006). Indigenous peoples get tired of telling and re-telling their history, their story. When you engage, after having done your homework, you display a genuine willingness and commitment to authentically engage with Indigenous people.

Second, look to expand your Indigenous networks. Start with people you know, and think broadly. It could be through the arts, sport, or a community project or organization. It may not be the specific Indigenous people you will engage with later, but you're building your "general" Indigenous cultural knowledge and understanding, which will help you in future engagements.

Third, undertake relevant professional development. This would include Indigenous-specific evaluation workshops as well as professional development in related fields, for example, culturally responsive evaluation, or workshops on equity, diversity, and combatting racism. You might also need to look to non-evaluation fields such as community development, social justice, the arts, or the environment—where they offer insights about working with Indigenous peoples, as well as Indigenous studies short courses or papers offered by tertiary education providers.

Elevating Indigenous Values and Principles in the Practice of Evaluation in Canada

In Aotearoa Maori, values and principles have been elevated in the practice of evaluation by the national association. ANZEA included the principles of the Treaty of Waitangi—partnership, protection, and participation—as part of its constitution. The inclusion of the Treaty principles establishes the uniqueness of evaluation in Aotearoa and focuses our evaluation leadership on what is needed to ensure the inclusion and participation of Indigenous peoples and Indigenous perspectives in all aspects of evaluation (Wehipeihana, Bailey, Davidson, & McKegg, 2014). It also sends a clear message to organizations that fund evaluation about the centrality of Māori values and principles as integral to the practice of evaluation in Aotearoa. This was evident in the development of the New Zealand Evaluator Competencies and Evaluation Standards and the integration of Māori perspectives into these professional practice publications in Aotearoa.

Understanding the Importance of Relationships when Engaging with Indigenous Peoples

In Māori and Indigenous contexts, relationships are the glue, the anchor, and the currency for effective and respectful engagement (Wilson, 2008)—and for successful evaluations. It is through relationships that the evaluation process unfolds. And it is in relationships, and in doing things together, that relational trust is built. Relational trust paves the way for deeper and more meaningful connections.

Relationships therefore are not something the evaluator simply pays attention to; they are inextricably linked to engaging with Indigenous people and therefore Indigenous evaluation (Wehipeihana et al., 2015).

Advocate for the Credibility and Inclusion of Indigenous Values, Methods, and Criteria

While there have been gains at a governmental level, which specifically require the use or consideration of Kaupapa Māori principles in evaluation involving Māori, significant challenges remain. The role of Māori as leaders, or as core members of evaluation teams, and the observance of cultural protocols are relatively well established. However, challenges remain in the determination of what counts as credible evidence, accepted methods, and the inclusion of Māori cultural values and principles in the framing of evaluative criteria.

Culturally grounded, national Māori health frameworks accepted and promoted by the New Zealand Ministry of Health (Ministry of Health, 2015—*Te Whare Tapa Whā* (The Four Cornerstones of Health) (Durie, 1994), *Te Pae Mahutonga* (Southern Cross Constellation—four elements of health promotion) (Durie, 1999), and *Te Wheke* (The Octopus—defining family health) (Pere, 1982)—have helped to elevate the importance of *tikanga Māori* (Māori cultural values) as critical to the assessment of Māori health outcomes. Furthermore, in 2019, there are positive indications of Māori values and principles being woven into national accountability frameworks for all New Zealanders, including Māori: “and that wellbeing considered from an Indigenous perspective moves the public policy discourse beyond Western constructs of wellbeing and enables an improved lived experience of wellbeing for everyone” (Te Puni Kōkiri & Treasury, 2019, p. i).

The Indigenous Approaches to the Living Standards framework developed by Te Puni Kōkiri (the Ministry for Māori Development) and the Treasury, while having a focus on wellbeing for Māori specifically, articulates a way of looking at wellbeing that can be applied to the full range of populations within Aotearoa New Zealand (see Box 1). The framework offers a way of accounting for various values and beliefs that drive people's experiences of wellbeing and positions the New Zealand public sector to advance wellbeing in a different way, looking to respond to the various needs, interests, and aspirations of all New Zealanders (Te Puni Kōkiri & Treasury, 2019).

Box 1. An Indigenous approach to the living standards framework

There is no one way to look at wellbeing. People view wellbeing differently depending on their values, beliefs and social norms. The way Māori view wellbeing is different from the way other New Zealanders view wellbeing. It is informed by *te ao Māori* (a Māori worldview) where, for example, *whenua* (land) is not seen just for its economic potential, but through

familial and spiritual connections defined by cultural concepts such as whakapapa (genealogy) and kaitiakitanga (stewardship). A te ao Māori perspective of wellbeing is also informed by life experiences—similar to that of other indigenous populations across the globe—of significant disparity and inequitable access to the tools, resources and opportunities that form the foundation to wellbeing. Te Tiriti o Waitangi, Aotearoa New Zealand's founding document, puts significant weight on partnership, active protection of Māori interests and redress to address past wrongs—including ongoing disparity and inequity experienced by Māori and their ability to access and benefit from capital stocks in various forms. When taken together they convey an obligation on the Crown and Māori to work together. To do this, the Crown—Ministers, departments and other agencies—must seek to understand te ao Māori, particularly as it relates to improving the wellbeing of whānau now, and over generations to come. Fortunately, te ao Māori offers a way to consider wellbeing within a holistic, robust and long-standing system. (Te Puni Kōkiri & Treasury, 2019)

Māori have long advocated that what's good for Māori is good for all New Zealanders, and the Indigenous Approaches to Living Standards framework is one example of culturally grounded, Indigenous principles and perspectives embedded within a national evaluative framework. The key message for evaluators is to promote and advocate for the inclusion of Indigenous values in evaluation with Indigenous peoples, and to argue the benefits for non-Indigenous peoples. It will also be important to persevere with these efforts because, based on the New Zealand experience, challenging and changing deep-seated values does not happen at pace, nor does it typically come about through all-encompassing, sweeping reform. In our experience, changing fundamental values comes about through uncompromising, incremental radicalism.

A PARADIGM SHIFT IS NEEDED

Non-Indigenous evaluators occupy a privileged position that has conferred the authority and power to define reality, to make impactful judgments about others, and for those judgements to be seen as accurate and valid (Johnson, 2001; Kirkhart, 2015; Sanakar, 2017). It can therefore be challenging for non-Indigenous evaluators to change their practice and to relinquish power and privilege. They have to want to do things differently or have a reason to see the world through alternative eyes. This may be for social justice reasons (Greene, 2011; Mertens, 2009), because of an ethic of care, or to do no harm, or they may be convinced by methodological or practice arguments such as multicultural validity (Kirkhart, 2010). Whatever their motivation, this is not simply about what they know and how they do evaluation; it is fundamentally about how they view the world—so a paradigm shift, or shifts will be needed. For example, this might include moving from ...

... evaluation as transactional to evaluation as relational

Evaluation as transactional is about a focus on the processes of evaluation, the methods, tools, and timeframes needed to undertake an evaluation. [Greene \(2005\)](#) reminds us that evaluation is about not only what we do but also who we are and where we position ourselves in relation to others. The paradigm shift, therefore, to evaluation as relational recognizes that relationships are the primary currency of engagement for Indigenous peoples ([Wehipeihana, 2013](#)) and sees evaluation as inherently relational.

... evaluator as expert to Indigenous people as experts

Evaluator as expert acknowledges evaluators as having typically Western-framed evaluation qualifications, expertise, and experience ([Wehipeihana & McKegg, 2018](#)). The paradigm shift to Indigenous peoples as expert affirms the unique cultural knowledge of Indigenous people to enter, navigate, and engage with their people ([Durie, 2001](#); [Smith, 1999](#)) and demonstrates that the currency of Western knowledge and experience diminishes in Indigenous contexts ([Wehipeihana, 2013](#)).

... evaluation as independent to evaluation as connected

Evaluation as independent and non-biased is about the privileging of Western methods, or the imposition of relation boundaries providing validity and credibility in evaluative judgments. The paradigm shift to evaluation as connected affirms being Indigenous, of the community, and a known face ([Cram, 2001](#); [Smith, 1999](#)) as central to the accuracy, credibility, and cultural validity of evaluative judgements ([Wehipeihana & McKegg, 2018](#)).

... evaluation managed by non-Indigenous evaluators to evaluation managed by Indigenous evaluators

When evaluations in Indigenous communities are managed by non-Indigenous evaluators, Indigenous evaluators are not always present when key documents are developed and agreed and when significant decisions are made. Even when Indigenous evaluators are present, they are often out-numbered, junior in status, or in environments where there is no willingness or appetite to consider alternative options or views. So the underlying issue is one of power and control over decision making in evaluation ([Wehipeihana et al., 2010](#)). When evaluations are managed by Indigenous evaluators, they (mostly) get to decide what's important and therefore prioritized, how resources are allocated, the criteria for evaluative judgements, what counts as valid evidence, and how findings are reported.

TOWARD INDIGENOUS CONTROL OF INDIGENOUS EVALUATION

Implicit in all of the paradigm shifts is the ceding of control or, as a minimum, the sharing of power and authority with Indigenous peoples. I developed a model⁴ ([Wehipeihana, 2013](#)) to explore to what extent evaluators share power and decision making in evaluation with Indigenous peoples.

Each aspect of the model is briefly discussed in Table 3. The model invites evaluators to reflect on their evaluation practice as a first step toward increasing decision making by Indigenous peoples in evaluation and with a clear goal of contributing to Indigenous-led evaluation. When applied as a self-assessment tool, it identifies how power is shared, and the extent to which Indigenous people have control over decision making in evaluation is made evident.

Some initial reflection questions were developed to support a self-assessment process:

- Who initiated the evaluation, and who benefits?
- What difference will it make for Indigenous peoples?
- Who will carry it out, and are they a good/appropriate “fit” for the community?
- How will the evaluation be carried out, and is the methodology a good/appropriate “fit” for the community?
- What is (or will be) the role of the Indigenous people or communities in the evaluation? Is this role resourced? How do they feel about their role in the evaluation? Does it match their expectations? And if not, can it be changed?
- What are the opportunities for you or the evaluation to move up and along on the continuum?
- What needs to happen to support this movement?

Figure 1. Wehipeihana Model 2013

However, reflective questions—with an Indigenous interrogation lens—can be asked at any or every step of the evaluation process, from the commissioning of the evaluation (e.g., who initiated the evaluation and was the Indigenous community consulted?); managing the evaluation (e.g., who are the stakeholders, are Indigenous people represented, and are decision-making processes equitable?); the design of the evaluation (e.g., are Indigenous principles and methods included in the evaluation design and are Indigenous evaluators, and ideally local people, part of the evaluation team?); the analysis of the data (e.g., is there a process for checking the accuracy and cultural validity of data-analysis and evaluative conclusions?); and dissemination (e.g., how will the evaluation findings be shared with Indigenous communities?).

CONCLUSION

This paper has advocated for Indigenous-led as being central to Indigenous evaluation. It argues for personal paradigm shifts by non-Indigenous evaluators and funders of evaluation as necessary to disrupt their taken-for-granted assumptions of control and to radically shift the power balance by placing control in the hands of Indigenous peoples. Decolonizing one’s evaluation practice is a deeply personal experience. It requires introspection, looking deeply at one’s values and beliefs, soul searching, and surfacing and interrogating implicit assumptions and biases. It also requires humility, being open to other perspectives and worldviews, to recognize unearned privilege, and to share, and ultimately let go of, power. And it calls for

Table 3. Description of the [Wehipeihana \(2013\)](#) model elements

Decision making	The horizontal axis is a decision-making and control continuum moving from Western control (you decide) through to Indigenous control (I decide).
Consequences	The vertical axis is a consequences continuum moving from good results and positive outcomes to no change, ineffective results, or harm.
Evaluation done to	Indigenous peoples have no say or control over the evaluation and there are no clear benefits for Indigenous peoples. Evaluation is done to Indigenous peoples, and Western worldviews and Western science prevail.
Evaluation done for	Seeks to improve the situation for Indigenous peoples but there is minimal consultation with Indigenous peoples. The evaluation is designed and managed without reference to Indigenous values, principles, and priorities, and decision-making power resides with the evaluator. Evaluation is paternalistic, and Western worldviews and Western science prevail.
Evaluation done with	Power and decision making are shared and negotiated. It is a collaborative and participatory space. Indigenous and Western worldviews and approaches are utilized.

(Continued)

Table 3. Continued

Evaluation done by	Indigenous people have control over the evaluation, and Indigenous methods and worldviews prevail. Evaluations may utilize Western methods.
Evaluation done as	Indigenous peoples have control over the evaluation, and Indigenous knowledge and science are the norm. The legitimacy and validity of Indigenous principles, values are taken for granted. It does not exclude Western methods but includes them only as far as they are seen to be useful.
Invitational space	The concept of the invitational space is where the evaluation is controlled by Indigenous people, and non-Indigenous evaluators accept that their participation is by invitation only.

courage, to embark on a journey that will be personally and professionally confronting, and tenacity, to stay the course and stand up to cynics, critics, and racists. It's an important first step.

At the same time, there is a need to confront institutional and structural barriers by supporting the development of Indigenous evaluators and an Indigenous evaluation community; elevating Indigenous values and principles in Canadian evaluation practice; advocating for the credibility and inclusion of Indigenous values, methods, and criteria within evaluation; and building the cultural competency of non-Indigenous evaluators and funders of evaluation.

NOTES

- 1 New Zealanders of European descent.
- 2 The *Tiaki* (Mentor) model, where research is guided by authoritative Māori; the *Whangai* (Adoption) model, where researchers become one of the *whānau*; the Power Sharing model, where the researcher and community jointly engage in the research; and the Empowering Outcomes model, where researchers provide information and answers to questions or topics that Māori want to find out about.
- 3 Evaluators who identify as Māori on the Ma te Rae Māori evaluation association membership list as of November 2018.
- 4 The model was initially titled "A framework for increasing control by Indigenous communities" for the AES 2013 Conference [Wehipeihana \(2013\)](#). Soon after (and also in 2013) I referred to it as "Locating evaluation practice: evaluation as an expression of power, control, and consequences." For the CES 2018 keynote presentation, I referred to it as "A framework for assessing evaluator progress towards Indigenous cultural competency and Indigenous-led evaluation." I tend to change the title to convey a key message or purpose to suit the audience or context.

REFERENCES

- [Bishop, R. \(2005\)](#). Freeing ourselves from neocolonial domination in research: A Kaupapa Maori approach to creating knowledge. In N. Denzin & Y. Lincoln (Eds.), *The SAGE handbook of qualitative research* (3rd ed., pp. 109–138). Thousand Oaks, CA: SAGE.

- Brayboy, B. M., & Deyhle, D. (2000). Insider-outsider: Researchers in American Indian communities. *Theory into Practice*, 39(3), 163–169. https://doi.org/10.1207/s15430421tip3903_7
- Chilisa, B. (2012). *Indigenous research methodologies*. Los Angeles, CA: SAGE.
- Cram, F. (1997). Developing partnerships in research: Pākehā researchers and Māori research. *SITES*, 35, 44–63.
- Cram, F. (2001). Rangahau Maori: Tona Tika, Tona Pono—The validity and integrity of Māori. In M. Tolich (Ed.), *Research ethics in Aotearoa New Zealand* (pp. 35–52). Auckland, New Zealand: Pearson Education.
- Cram, F. (2009). Maintaining indigenous voices. In D. Mertens & P. Ginsberg (Eds.), *The SAGE handbook of social science research ethics* (pp. 308–322). Thousand Oaks, CA: SAGE.
- Cram, F. (2016). Lessons on decolonizing evaluation from Kaupapa Māori evaluation. *Canadian Journal of Program Evaluation*, 30(3), 296–312. <https://doi.org/10.3138/cjpe.30.3.04>
- Cram, F., Chilisa, B., & Mertens, D. M. (2013). The journey begins. In D. M. Mertens, F. Cram, & B. Chilisa (Eds.), *Indigenous pathways into social research: Voices of a new generation* (pp. 11–40). Walnut Creek, CA: Left Coast Press.
- Durie, M. (1994). *Whaiora: Māori health development*. Auckland: Oxford University Press.
- Durie, M. (1999, December). Te Pae Mahutonga: A model for Māori health promotion. *Health Promotion Forum of New Zealand Newsletter*, 49, 2–5.
- Durie, M. (2001). *A framework for considering Māori educational advancement: The Hui Taumata Matauranga*. Taupo, New Zealand: Massey University.
- 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. <https://doi.org/10.18296/em.0003>
- Greene, J. C. (2005). Evaluators as stewards of the public good. In S. Hood, H. Frierson, & R. Hopson (Eds.), *The role of culture and cultural context in evaluation: A mandate for inclusion, the discovery of truth and understanding* (pp. 7–20). Charlotte, NC: Information Age Publishing.
- 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.
- Harvey, G. (2003). Guesthood as ethical decolonising research method. *Numen: International Review for the History of Religions*, 50(2), 125–146. <https://doi.org/10.1163/156852703321506132>
- Health Research Council of New Zealand. (2010). *Guidelines for research on health research involving Māori*. Retrieved from <http://www.hrc.govt.nz/news-and-publications/publications/guidelines-researchers-health-research-involving-m%C4%81ori>
- Henare, M. (1988). Ngā Tikanga Me Ngā Ritenga O Te Ao Māori. Standards and foundations of Māori society. In *The April report: Report of the Royal Commission on Social Policy Future; Directions. Vol 3, part 1: Future directions, associated papers* (pp. 3–42). Wellington, New Zealand: The Royal Commission on Social Policy.

- Johnson, A. (2001). *Privilege, power and difference*. Mountain View, CA: Mayfield.
- Kirkhart, K. E. (2010). Eyes on the prize: Multicultural validity and evaluation theory. *American Journal of Evaluation*, 31(3), 400–413. <https://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. <https://doi.org/10.18296/em.0002>
- LaFrance, J., & Nichols, R. (2010). Reframing evaluation: Defining an indigenous evaluation framework. *Canadian Journal of Program Evaluation*, 23(2), 13–31.
- Marsden, M., & Royal, T. A. C. (2003). *The woven universe: Selected writings of Rev. Māori Marsden*. N.p.: Estate of Rev. Māori Marsden.
- Mertens, D. M. (2009). *Transformative research and evaluation*. New York, NY: Guilford Press.
- Ministry of Health. (2015). *Māori health models*. Retrieved from <https://www.health.govt.nz/our-work/populations/maori-health/maori-health-models>
- Ministry for Social Development. (2004). *Ngā Ara Tohutohu Rangahau Māori—Guidelines for Research and Evaluation with Māori*. Centre for Social Research and Evaluation Te Pokapū Rangahau Arotake Hapori. Wellington, New Zealand: Ministry for Social Development.
- Ormond, A., Cram, F., & Carter, L. (2006). Researching our relations: Reflections on ethics. *Alternative: An International Journal of Indigenous Scholarship*, Special supplement 2006—Marginalisation, 180–198.
- Pere, R. (1982). *Ako: Concepts and Learning in the Maori tradition*. Hamilton, New Zealand: University of Waikato Press.
- Sanakar, M. (2017). *A stocktake on culturally responsive evaluation in and outside the UN system*. A discussion document prepared for UNEG Strategic Objective 3: Evaluation informs UN system-wide initiatives and emerging demands. United Nations Evaluation Group, UNESCO Evaluation Office.
- Smith, G. H. (1990). *Research issues related to Māori education*. Paper presented at the NZARE Special Interests Conference.
- Smith, G. H. (1997). *The development of Kaupapa Māori: Theory and praxis* (Unpublished doctoral thesis). University of Auckland, Auckland, New Zealand.
- Smith, L. T. (1999). *Decolonising methodologies: Research and Indigenous peoples*. Dunedin, New Zealand: University of Otago Press.
- Spiller, C., & Stockdale, M. (2013). Managing and leading from a Māori perspective: Bringing new life and energy to organisations. In J. Neal (Ed.), *Handbook of faith and spirituality in the workplace: Emerging research and practice* (pp. 149–173). New York, NY: Springer. https://doi.org/10.1007/978-1-4614-5233-1_11
- Teariki, C., Spoonley, P., & Tomoana, N. (1992). *Te Whakapakari te Mana Tangata: The Politics and Process of Research for Māori*. Palmerston North, New Zealand: Massey University.
- Te Awakotuku, N. (1991). *He tikanga whakaaro: Research ethics in Māori community*. A discussion paper. Wellington, New Zealand.

- Te Puni Kōkiri. (1999). *Evaluation for Māori: Guidelines for government agencies*. Aroturuki me te Arotakenga Monitoring and Evaluation Branch. Wellington, New Zealand: Te Puni Kōkiri.
- Te Puni Kōkiri and Treasury. (2019). *An Indigenous approach to the living standards framework*. The Treasury Discussion Paper 19/01. Wellington, New Zealand: Treasury.
- Walker, H. (1990). *Māori attitudes on research*. Wellington, New Zealand: Department of Social Welfare
- Wehipeihana, N. (2013). *A vision for Indigenous evaluation*. Keynote paper presented at the Australasian Evaluation Society Conference.
- Wehipeihana, N., Bailey, R., Davidson, E. J., & McKegg, K. (2014). Evaluator competencies: The Aotearoa New Zealand experience. *Canadian Journal of Program Evaluation*, 28(3), 49–69.
- Wehipeihana, N., Davidson, E. J., McKegg, K., & Shanker, V. (2010). What does it take to do evaluation in communities and cultural contexts other than our own? *Journal of MultiDisciplinary Evaluation*, 6(13), 182–192.
- Wehipeihana, N., & McKegg, K. (2018). Values and culture in evaluative thinking: Insights from Aotearoa New Zealand. *New Directions for Evaluation*, 158, 93–107. <https://doi.org/10.1002/ev.20320>
- Wehipeihana, N., McKegg, K., Thompson, V., & Pipi, K. (2015). Cultural responsiveness through developmental evaluation: Indigenous innovations in sport and traditional Māori recreation. In M. Q. Patton, K. McKegg, & N. Wehipeihana (Eds.), *Developmental evaluation exemplars: Principles in practice* (pp. 25–44). New York, NY: Guilford Press.
- Wilson, S. (2008). *Research is ceremony: Indigenous research methods*. Black Point, NS: Fernwood.

AUTHOR INFORMATION

Nan Wehipeihana Nan is an independent evaluation consultant, based in Wellington New Zealand. Nan has more than 20 years evaluation experience and specializes in evaluation and research with a focus on Māori (the Indigenous people of New Zealand). Nan established Research Evaluation Consultancy in 1997 and is a member of the Kinnect Group. Nan is a former board member of the Aotearoa New Zealand Evaluation Association (ANZEA) and a former appointed executive member of the Australasian Evaluation Society (AES). Nan's iwi (tribal) affiliations are to Ngāti Porou and Te Whānau-ā-Apanui on the East Coast of New Zealand and Ngāti Tukorehe and Ngāti Raukawa, north of Wellington, New Zealand.

Book Reviews / Comptes rendus de livres

Anne Vo and Christina A. Christie (Eds.). (2015). *Evaluation Use and Decision Making in Society: A Tribute to Marvin C. Alkin*. Charlotte, NC: Information Age Publishing. Paperback, 190 pages. (ISBN: 978-1-68123-004-7)

**Reviewed by Margaret Schultz Patel, University of Denver/
Schultz Patel Evaluation**

Compared to other branches of research, the field of evaluation is uniquely concerned with producing actionable information. Without this charge, it is not clear what would distinguish evaluation from other types of applied research. Despite this seemingly simple focus, seasoned evaluators know all too well that there are often many barriers to shepherding evaluations toward achieving this final purpose. The volume under review contains the musings of seasoned evaluators and evaluation theorists grappling with these barriers from many different perspectives.

This book, which is the result of a symposium dedicated to celebrating the career of evaluation theorist Marvin C. Alkin, is a compilation of meditations about evaluation use from various scholars and thought leaders in the field. Evaluation use can be defined as the extent to which evaluation findings are used to inform practice and policy. These reflections are based primarily on the authors' lived experience as evaluators and evaluation theorists. One thing that becomes clear early on is that what we know about evaluation use—empirically speaking—is pretty paltry. Considering the youth of evaluation as a field, as well as the challenge of consistently operationalizing a latent construct such as use, this is not surprising.

To frame the subsequent discussion of evaluation use and decision making, the book begins with a clear discussion of the related historical context by Christina Christie, who reminds us that “evaluation grew out of the need to better understand the impact of U.S. federal policies” (p. 1). An allegory of how to facilitate decision making with soundly reasoned evaluation by Ernie House provides a tangible example for readers. A convincing argument that evaluations informed by strong evaluation theory are more likely to be used for decision making and why—hint: it has to do with the strength of evidence generated by theory-informed evaluation—is subsequently articulated by Stewart Donaldson and Tarek Azzam.

Next, Eric Barela shares his experiences as an internal evaluator grappling with the various tensions in balancing rigor with relevance and compliance with

Corresponding Author : Margaret Schultz Patel, maggie.patel@du.edu

learning. Barela integrates Volkov's (2011) conceptual framework of tensions inherent to the role of the internal evaluator with the well-known taxonomy of evaluation theories (Alkin & House, 1992). This application of an important theoretical framework to a pragmatic assessment of the roles internal evaluators play felt to this reader like a rallying cry—there is a way to meaningfully grapple with the tensions experienced as an internal evaluator, and trying to bulldoze your colleagues with a one-size-fits-all approach is not the best way to encourage evaluation use. When compared with their external counterparts, internal evaluators are arguably best positioned to influence an organization to use evaluation to guide decision making over the long term, and this chapter will serve as an invaluable decision tree for those grappling with how to do so.

Richard Nunneley, Jean King, Kelli Johnson, and Laura Pejsa contribute a literature review of the empirical evaluation literature on “how use and influence affect decision making” (p. 55). In an elegantly argued and largely philosophical piece, they demonstrate how little we actually know about these processes and raise a call for more direct assessments of this question, as “rigorous study, theory development, theory testing, and healthy debate around theory are not simply learning exercises, but an investment in our field and its potential impact on our society” (p. 68). Eleanor Chelimsky, a long-time government evaluator, shares her experiences from the field related to barriers to evaluation use in that context. These hard-earned lessons, such as understanding when an evaluation has been commissioned for purely symbolic reasons, should be considered by proactive evaluators working to get ahead of the curve on barriers to evaluation use. The next two articles argue convincingly for the importance of incorporating additional perspectives into efforts to ensure evaluation use: accountability and learning through a transformative lens (J. Bradley Cousins, Katherine Hay, and Jill Chouinard) and culturally responsive decisions in evaluation practice (Wanda Casillas, Rodney Hopson, and Ricardo Gómez). Both chapters contain important practical guidance that is useful for helping the field move beyond the same old challenges that seem to hamper evaluation use. Michael Quinn Patton, pioneer of utilization-focused evaluation, then explicates a convincingly argued diatribe on what may lead to evaluations being misused (hint: poorly-done evaluations factor prominently, though incompetence is one of the more innocuous drivers of misuse in Patton's estimation). Editor Anne Vo concludes the volume with a helpful synthesis of what was learned and what is needed to drive the field forward.

This book is a great read when one is in need of some refocusing or revving up one's passion for evaluation. This book will be a treat for workaday practitioners trapped in the relentless cycle of responding to Requests for Proposals (RFPs), tempering client expectations, or paring down their own evaluative aspirations. Perhaps more importantly, this volume picks up a conversation in the literature about how good evaluation is at achieving what many would consider its core mission: influencing decision making. This is an idea that should be of supreme importance to all evaluation practitioners and academic researchers of evaluation.

This volume highlights many ways to think about use and how to combat lack of use or misuse. As a whole, it does a great job summarizing what we know, empirically and anecdotally, about evaluation use. Beyond this, the volume also contains a charge to get better at delivering relevant and actionable evaluation findings. This book is an important contribution to a field still trying to define itself and realize its potential.

REFERENCES

- Alkin, M., & House, E. (1992). Evaluation of programs. In M. Alkin (Ed.), *Encyclopedia of educational research* (6th ed., pp. 462–467). New York, NY: Macmillan.
- Volkov, B. B. (2011). Beyond being an evaluator: The multiplicity of roles of the internal evaluator. In B. B. Volkov & M. E. Baron (Eds.), *Internal evaluation in the 21st century. New Directions for Evaluation*, 132, 25–42.

Book Reviews / Comptes rendus de livres

Scott G. Chaplowe and J. Bradley Cousins. (2016). *Monitoring and Evaluation Training: A Systematic Approach*. Thousand Oaks, CA: SAGE. Hardcover, 439 pages. (ISBN 978-1-4522-8891-8)

**Reviewed by Michele Tarsilla, UNICEF, West and Central Africa,
Regional Office, Dakar, Senegal**

When I first looked at the title of this book, I confess that I had mixed feelings. First, the book's focus on "monitoring and evaluation" seemed a bit too broad: monitoring and evaluation certainly complement each other; however, their respective specificities risk being neglected when the "M" and the "E" are bundled together. Next, the book's focus on "training" seemed too narrow to tackle the complexity of the M&E capacity realms. Specialized literature confirms that training by itself is only one of the many strategies one could pursue to strengthen M&E capacity. I have argued for years that the excessive attention given to training (a phenomenon that I refer to as the objectification of capacity) has undermined efforts made thus far to strengthen national M&E capacity. To the contrary, it is through interventions and strategies geared toward capacity strengthening at not only the individual but also the organizational and systemic levels that sustainable and contextually relevant Evaluation Capacity Development (ECD) occurs (Tarsilla, 2014).

Two factors motivated me to delve into this book despite an initial skepticism. First was my familiarity with and appreciation of the co-author's prior work (Bradley Cousins is a professor at the University of Ottawa and a prolific author of articles on capacity building and participatory approaches; Scott Chaplowe is a senior evaluation specialist with an extensive range of experience both within and outside of numerous international organizations). Second, the "systematic approach" referred to in the subtitle set this resource apart from the most M&E training manuals available on the market, which often jump directly into Kirkpatrick's (1996) model and a few facilitation tricks to get workshop participants learning about M&E.

Fortunately, some of my earlier doubts dissipated rapidly once I started reading the book. Right from the introduction, I was positively surprised by the authors' continued reference to the need for "systemic thinking" when planning, designing, and developing training events. Likewise, I enjoyed the insightful presentation of the adult learning principles, different types of learning, and key trainers' competencies. Page after page, the book made me reflect upon my own

Corresponding Author: Michele Tarsilla, mtarsilla@unicef.org

practice and spurred some new ideas about how I could enhance the quality of my training planning in the future. However, in doing so, I also realized that the book is indeed quite dense (a total of 405 pages distributed across 11 chapters, and one supplementary section including 99 real-life examples of capacity building) and that the less-experienced reader may feel a bit intimidated by the vast quantity of information provided. That notwithstanding, the book is very well written (I read it cover to cover over a weekend) and logically structured: each chapter starts with clearly articulated objectives and key definitions (“learning transfer” and “utility” are explained quite cogently) and ends with a critical summary and a list of recommended technical resources for future consultation.

The book is divided into three parts. Part 1 (“Key Concepts for M&E Training”) provides readers with an understanding of the foundational elements of M&E training and clarifies different elements and types of impactful training (from formal to non-formal training; from informal to incidental). I especially appreciated the focus of Chapter 1 on how to enhance training usefulness, often referred to as “training transfer,” as well as of Chapter 3, dedicated to the need to design contextually relevant training programs, as training is not isolated from, and requires a synchronization with, trainees’ organizational culture as well as the external environment with which those organizations interact. Likewise, I found Chapter 4 on adult learning particularly instructive: it provided a solid theoretical grounding for some of the most influential conceptual frameworks in the training field: from Knowles’s six assumptions of learning to andragogy and the taxonomy of learning styles (i.e., activists, reflectors, theorists, and pragmatists). Chapter 5 is also quite unique in that it discusses the six trainer key competencies. The only downside is that the book as a whole relates mostly to mainstream M&E and not does not explain how training could be adapted when used in relatively new evaluation domains, such as developmental evaluation, or more complex environments, such as public policy evaluation training.

Part 2 (“A Systematic Approach to M&E Training”) is more practical in scope and spells out the different phases of the ADDIE framework (training analysis, design, development, implementation, and evaluation). The checklist at the end of Chapter 6 on actions required to be taken during each phase of the framework is particularly interesting, as it helps give structure to a process that, due to the large number of parties involved, is often scattered and/or not well sequenced. Chapter 7 on needs and outcomes analysis is quite useful too, as it suggests the use not only of “backward design” (Wiggins & McTighe, 2005) during the identification and verification of training outcomes, but also of trainee analysis coupled with task and gap analysis, as a way to better enhance the training utility. Likewise, I found Chapter 8 on the selection of learning objectives quite interesting, especially the section discussing the necessary linkages among performance, condition, and criterion, although it is rather detailed in places. In real circumstances, a trainer would hardly include in a training plan all of the following: training outcomes, learning goal, primary learning objective, and secondary learning objectives. Furthermore, Chapter 10 stands out for its provision

of tips on how to handle and successfully eliminate disruptive behaviour during the training time. Lastly, Chapter 11 skillfully covers the different approaches to evaluating training programs.

Part 3 (“M&E Training Methods and Techniques”) presents the reader with a plethora of techniques (99 in total) that are expected to enrich the quality and effectiveness of M&E training (e.g., from learning games and practicum experiences to role-playing and M&E software activities). Details for each of these tools, including a balanced discussion of related advantages and disadvantages, are particularly useful; however, the authors could have discussed in greater depth the implications of the use of some of these techniques during online training deliveries, as this section is focused mostly on face-to-face M&E training. Even with these caveats, and although some techniques replicate those found in other manuals, I am likely to adopt the more inventive exercises discussed, including popcorn introductions and group résumés, for helping break the ice among training participants and creating a bond among them from the beginning of the training program: the “if ... then” exercise for helping trainees develop more logical and evidence-based theories of change for their respective clients and organizations; the WEM3 exercise for getting trainees used to taking different—and somewhat diverging—perspectives into account during the planning, conduct, and dissemination of evaluations; and the participant-driven case studies and the fishbowl simulations for anchoring the training to the participants’ life practice and enhancing the quality of evaluation peer review.

Overall, this book represents an ideal addition to the “toolbox” of (i) training professionals with a sufficient knowledge of M&E concepts, (ii) M&E training managers, (iii) educators and students in post-secondary and continuing contexts, and (iv) training of trainers (ToT) facilitators. Despite not being able to tackle the issues of M&E capacity development at a more systemic level, this book has the merit of bringing rigour and intentionality back into the planning, design, and delivery of M&E training, while also providing a rich theoretical and sound experiential base.

REFERENCES

- Kirkpatrick, D. (1996). Revisiting Kirkpatrick’s four-level-model. *Training & Development*, 1, 54–57.
- Tarsilla, M. (2014). Evaluation capacity development in Africa: Current landscape of international partners’ initiatives, lessons learned and the way forward. *African Evaluation Journal*, 2(1), Art. #89, 13 pages. <https://doi.org/10.4102/aej.v2i1.89>
- Wiggins, G., & McTighe, J. (2005). *Understanding by design*. Alexandria, VA: Association for Supervision and Curriculum Development (ASCD).