A Framework to Combine Mixed Methods Integration and Developmental Evaluation to Study Complex Systems

Stephen MacGregor and Amanda Cooper
Queen’s University, Ontario, Canada

Abstract: This article considers the opportunities and challenges of mixed methods integration within the context of conducting developmental evaluations with diverse stakeholders. The article proposes a four-part planning framework that explicitly maps the eight principles of developmental evaluation to key questions about integration: Why integrate? What will be integrated? When and how will integration occur? The framework is used to explore and compare two large-scale evaluations that employed mixed methods integration. Ultimately, the work illustrates a need to balance data collection with the momentum of moving initiatives forward, check assumptions about stakeholder co-creation, and cultivate integration literacy.

Keywords: co-creation, complexity, developmental evaluation, mixed methods, research impact

Résumé: L’objectif principal de cet article est d’examiner les possibilités et défis de l’intégration de méthodes mixtes dans le contexte de la conduite d’évaluations dans une optique de développement avec diverses parties prenantes. L’article propose un cadre de planification en quatre parties pour les équipes de recherche: Pourquoi intégrer? Qu’est-ce qui sera intégré? Quand et comment l’intégration aura-t-elle lieu? Le cadre est utilisé pour explorer et comparer deux évaluations à grande échelle qui ont cherché à utiliser l’intégration des méthodes mixtes. En fin de compte, le travail montre la nécessité d’équilibrer la collecte de données et l’élan de faire avancer les initiatives, de vérifier les hypothèses sur la co-création des parties prenantes et la nécessité de cultiver la culture de l’intégration.

Mots clés: évaluation évolutive, méthodes mixtes, impact, complexité, co-création

Developmental evaluation (DE) is one of the most significant advances in the field of evaluation over the past three decades (Gamble, 2008). A departure from the traditional formative-summative distinction, DE supports the development of innovations within complex, dynamic systems (Patton, 2011). It specifically deals with “collaborative, complex, and evolving change processes” (Preskill &
Beer, 2012, p. 7) and the critical role of relational trust in generating learning and supporting the use of evaluative feedback (Patton, 2016a, 2016b). However, as evidenced in various practical guides (e.g., Baylor et al., 2019; Dozois et al., 2010; Gamble, 2008), effective practice in DE remains a challenge for many evaluators. This challenge partly stems from the “small and demanding” (Patton, 2011, p. 15) niche for DE, which limits its situational appropriateness, as well as its greater demands on evaluation competency domains such as interpersonal practice (Patton, 2016a). One area that remains without systematic guidance is the application of mixed methods (MM) within a DE context, despite the recent trend of evaluators working at the intersection of these domains (e.g., Guilcher et al., 2020; Laycock et al., 2019; Wu et al., 2018).

At the same time, it is important to bear in mind Patton’s (2011) specification that DE “is purpose- and relationship-driven not methods-driven” (p. 288). The appropriateness of any method must be determined in consideration of the evaluation purpose, context, participants, and intended uses. When it is appropriate given these factors, we have found the application of MM in DE to offer opportunities for understanding and responding to complex problem contexts in ways not possible through quantitative or qualitative methods alone. This potential emerges through integration, the backbone of MM, referring to how the quantitative and qualitative strands of an evaluation are interwoven (Creswell & Plano Clark, 2018). However, we have observed that what compounds the lack of guidance for MM in a DE context is a critical lack of exemplars for how integration unfolds. Without detailed examples and practical illustrations, evaluators are left with little guidance on where to focus their capacity-building efforts or how to support DE with the added complexity of MM effectively.

Hence, in this article, we interrogate the opportunities and challenges of bringing together DE and MM by drawing on two illustrative cases from the Research Informing Policy, Practice, and Leadership in Education (RIPPLE) program of research at Queen’s University. Examining units of analysis at the provincial and national levels, we offer considerations about the plausibility and value of bringing together DE and MM, what has worked or failed to work in practice, and technical guidance for evaluation practice as well as recommendations for evaluation research. The groundwork for this analysis begins with a review of the essential details of DE and MM, followed by an overview of our conceptual framework for charting their intersection.

WHAT IS DEVELOPMENTAL EVALUATION?

DE is a dynamic and responsive approach to evaluation that situates stakeholder engagement and co-creation as the primary mechanisms of change within complex systems. Through its orientation toward systems thinking and complex nonlinear dynamics, DE “supports innovation and adaptive management . . . [by] asking evaluative questions, applying evaluation logic, and gathering real-time data to inform ongoing decision making and adaptations” (Patton, 2011, p. 1). Patton (2016c) outlines eight essential principles that guide DE (Table 1).

doi: 10.3138/cjpe.71488
Table 1. Eight principles of developmental evaluation (summarized from Patton, 2016c)

<table>
<thead>
<tr>
<th>Principles of DE</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Developmental purpose</td>
<td>Five purposes for DE include (i) developing (creating, inventing) a new intervention aimed at a significant problem, (ii) ongoing adaptive development (based on changing conditions, new knowledge, or new clientele), (iii) developing greater impact by adapting validated innovation principles and practices to a new context (scaling), (iv) developing changes in and across systems, (v) developing rapid responses in crisis situations.</td>
</tr>
<tr>
<td>2. Evaluation rigour</td>
<td>“Ask probing evaluation questions; think and engage evaluatively; question assumptions; apply evaluation logic; use appropriate methods; and stay empirically grounded—that is, rigorously gather, interpret, and report data” (p. 296). Data anchors continuous and iterative learning cycles as initiatives progress, change, and grow in complex environments.</td>
</tr>
<tr>
<td>3. Utilization focus</td>
<td>“Focus on the intended use by intended users from beginning to end, facilitating the evaluation process to ensure utility and actual use” (p. 300). The focus on end-users is critical in DE, with stakeholder engagement intricately connected to evaluative processes longitudinally over the course of an initiative.</td>
</tr>
<tr>
<td>4. Innovation niche</td>
<td>“Elucidate how the change processes and results being evaluated involve innovation and adaptation, the niche of developmental evaluation” (p. 302). Innovation in this principle should be defined by the stakeholders rather than dictated by the evaluation team or funder. Engaging the perspectives of people on the front lines allows for adaptation and uptake that can actually help.</td>
</tr>
<tr>
<td>5. Complexity perspective</td>
<td>“Understand and interpret development through the lens of complexity and conduct the evaluation accordingly. This means using complexity premises and dynamics to make sense of the problems being addressed; to guide innovation, adaptation, and systems change strategies; to interpret what is developed; to adapt the evaluation design as needed; and to analyze emergent findings” (p. 304). Complexity theory brings ideas of nonlinearity, emergence, adaptation, and uncertainty into the mix. Tensions and disagreements are inevitable in most partnerships, and societal systems are marked by their complexity (e.g., the number of organizations, roles, funders, interactions with other sectors, community attributes, regional differences in context and cultures).</td>
</tr>
</tbody>
</table>

(Continued)
Table 1. (Continued)

<table>
<thead>
<tr>
<th>Principles of DE</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>6. Systems thinking</td>
<td>“Think systematically throughout, being attentive to interrelationships, perspectives, boundaries, and other key aspects of the social system and context within which the innovation is being developed and the evaluation is being conducted” (p. 306). By enacting systems thinking, DE teams in conjunction with stakeholders can look at how different parts of the system interact and draw conclusions about how these patterns of interaction influence the system.</td>
</tr>
<tr>
<td>7. Co-creation</td>
<td>“Develop the innovation and evaluation together—interwoven, interdependent, iterative, and co-created—so that developmental evaluation becomes part of the change process” (p. 306). This is potentially one of the most important aspects of DE.</td>
</tr>
<tr>
<td>8. Timely feedback</td>
<td>“Timely feedback to inform ongoing adaptation as needs, findings, and insights emerge, rather than only at predetermined times (e.g. quarterly, or at midterm of the end of project)” (p. 308). Perhaps the most challenging aspect of DE is the speed of change and the narrow windows of opportunity to leverage change within a particular space, time, and stakeholder group. Utility, however, remains at the heart of DE, and if the stakeholders cannot use the data (which requires timeliness and targeted recommendations), then a major benefit of the DE approach is squandered.</td>
</tr>
</tbody>
</table>

As Cousins and Shulha (2006) note, “possibly the most significant development of the past decade in both research and evaluation communities has been a more general acceptance that how we work with clients and practitioners can be as meaningful and consequential as what we learn from our methods” (p. 277). Accordingly, a DE team often takes on a collaborative role in tracing impact across systems; rather than adopting a neutral external perspective from afar, evaluators are embedded in change processes and their primary role is to ask questions, use data to reflect the system at a given time, ask stakeholders what they think might explain the patterns, and brainstorm collective solutions or avenues to improved processes or impacts. If done well, the DE itself becomes part of the innovation as initiatives unfold. In any DE, all of these principles are present. However, as Patton (2016b) advises, evaluators must interpret and apply these principles within the specific situation and context, not like items on a checklist.

The connection with MM emerges from the openness of DE to employ the combination of methods most likely to promote the evaluation’s intended use(s) by the intended users: DE “does not rely on or advocate any particular evaluation method, design, tool, or inquiry framework” (Patton, 2016a, p. 10). This openness ensures an ability to implement the methods appropriate for the problem context at hand, which may require MM and the integrated insights inaccessible by either qualitative or quantitative data alone.
WHAT IS MIXED METHODS?

Regularly described as the “third methodological movement” (Tashakkori & Teddlie, 2003, p. ix) in the social and behavioural sciences, MM has garnered increasing interest over the past 30 years as a bridge between the quantitative and qualitative research paradigms (Ghiara, 2020). The value of MM arises from it being “an intuitive way of doing research that is constantly being displayed throughout our everyday lives” (Creswell & Plano Clark, 2018, p. 1). In other words, MM can enable a more nuanced and comprehensive understanding of the myriad factors at play in complex problem contexts (McKim, 2017; Mertens, 2018; Poth, 2018), such as those that fit the niche and purpose of DE. Many definitions of MM exist (e.g., Johnson et al., 2007), invoking various philosophical (ontological, epistemological, axiological) and methodological positions (Maxwell, 2011). We adopt Creswell and Plano Clark’s (2018) definition, which frames MM¹ as a way of doing research/evaluation in which the researcher/evaluator

• “collects and analyzes both qualitative and quantitative data rigorously in response to research questions and hypotheses,
• integrates (or mixes or combines) the two forms of data and their results,
• organizes these procedures into specific research designs that provide the logic and procedures for conducting the study, and
• frames these procedures within theory and philosophy.” (p. 4)

As Creswell and Plano Clark go on to note, the crux of this definition is not just that both quantitative and qualitative methods are used; it is that they are integrated. Without integration, an evaluation would not merit the MM designation; more appropriately, it might be labelled multimethod (Anguera et al., 2018) or quasi-mixed methods (Teddlie & Tashakkori, 2009).

Despite its importance, integration is perhaps the least developed and most challenging aspect of MM (Creswell & Plano Clark, 2018; Tunarosa & Glynn, 2017). Early discussions about integration tended to focus on its role within the data-analysis process (Johnson & Onwuegbuzie, 2004) or suggest the general notion of a continuum from weakly integrated to fully integrated (Teddlie & Tashakkori, 2009). Fetters and Molina-Azorin (2017) define integration as an intentional “linking of qualitative and quantitative approaches and dimensions together to create a new whole or a more holistic understanding than achieved by either alone” (p. 293). It thus deals with the thinking and actions that enable the quantitative and qualitative data sources to interface in ways that can generate otherwise inaccessible insights (Poth, 2018). However, the specific procedures of integration vary according to the MM design employed (e.g., see Creswell & Plano Clark, 2018).

Recently, a number of MM scholars have advanced what it looks like to achieve meaningful integration in practice (cf. Fetters & Molina-Azorin, 2017; Lynam et al., 2020; Palinkas et al., 2019; Plano Clark, 2019). What makes this work
so promising is the apparent convergence on similar dimensions of integration. In particular, drawing upon the work of Plano Clark (2019) and Lynam et al. (2020), which offers a parsimonious framework, integration involves attention to the evaluation context in addition to why integration is planned, what will be integrated, and when and how integration will occur. As our focus in this article is MM within the scope of DE, we now turn to a discussion of how these questions take form within a DE context.

A FOUR-PART FRAMEWORK TO COMBINE MM AND DE

As set out above, DE and MM are independently complex domains of theory and practice. Their intersection is doubly so, providing a complex toolset with which evaluators can investigate how the components of social systems “interact in multiple, nonlinear ways without direction” (Poth, 2018, p. 5) to produce uncertain outcomes. In grappling with complexity, Poth (2018) underscores the need to eschew oversimplifications of MM approaches, both in research and evaluation. For that reason, and with the aim of providing actionable insights from the illustrative cases that follow, we introduce a novel framework for combining DE and MM (Figure 1) that is necessarily partial. Our framework seeks to promote integrity in DE by explicitly mapping the eight principles of DE to key questions about integration. It also provides general guidance for navigating the “zone of complexity” (see Patton, 2011): the problem space of DE, marked by high levels of uncertainty and social conflict about how to achieve desirable outcomes. The cross-case evaluation question we consider in this article is this: How was MM integration utilized in two DEs to understand complex network activities across diverse stakeholders?

**Context**

Our framework is based in the various contexts that are affected by key players in the DE. Context modulates all other dimensions, and maintaining an awareness of it respects the DE principles of systems thinking and embracing a complexity perspective. Increasingly conceptualized as process rather than place (May et al., 2016), context is less a backdrop for innovation and more a continually varying set of factors within which an innovation is embedded, “comprising not only a physical location but also roles, interactions and relationships at multiple levels” (Pfadenhauer et al., 2017, p. 6). Our framework engages with this complexity by encouraging the intentional, systematic exploration of nested contexts that are “infinitely complicated, intertwined and in motion” (Pawson, 2014, p. 37). In this way, and like the framework itself, developing an understanding of context is an always-in-motion process.

**Why integrate?**

The first dimension of our framework asks, “Why integrate?” This question is foremost reflected in the developmental purpose, utilization, and innovation niche principles of DE. Collectively, these principles portend a focus on the utility and
appropriateness of MM within the DE of a particular social innovation. In our experiences, developing an answer to this question in practice first requires clarity on the characteristics and circumstances of DE and MM within the innovation’s context(s) and with attention to stakeholders’ goals. With clarity established, a MM evaluation question that signals to the evaluation’s integrative nature must be co-constructed with stakeholders (Plano Clark, 2019). As Lynam et al. (2020) suggest, such questions tend to reference the ability of MM (via integration) to
generate “improved knowledge or understanding (breadth, depth, and validity) . . .
or] social change” (p. 342). The nature of the MM evaluation question varies ac-
cording to the objects being integrated.

What will be integrated?
The second dimension asks, “What will be integrated?” This question reflects the
evaluation rigour principle of DE, which concerns the application of evaluation
logic and the use of appropriate methods. The answer to this question should
make clear how integration is organized by “phases and stages but also issues such
as philosophical, theoretical and research integrity dimensions that permeate the
entire mixed methods research process” (Fetters & Molina-Azorin, 2017, p. 293).

As Fetters and Molina-Azorin (2017) illustrate, the objects of integration could
potentially include up to 15 dimensions (e.g., philosophical dimension, team di-
mension, sampling dimension). However, within the complex contexts of DE, it is
sometimes helpful to abbreviate this framework by three levels (see Fetters et al.,
2013): design (i.e., basic and complex MM designs), methods (i.e., connecting,
building, merging, and embedding quantitative and qualitative strands), and inter-
pretation and reporting (i.e., showcasing the rich complexity of MM). With the
level(s) of integration specified, the next task is how to make integration relational.

When and how will integration occur?
The final dimensions ask, “When and how will integration occur?” We have found
it useful to align the two parts of this question with the timely feedback and co-
creation principles of DE, respectively. Generating timely and ongoing feedback is
critical for DE to “detect how the context is changing, affirm current practices or
inform new activities, and guide programmatic and strategic questions regarding
the innovation’s progress” (Preskill & Beer, 2012, p. 7). We consider these dimen-
sions in relation to how opportunities for integration, as points of interface (Plano
Clark, 2019), enable single-loop and double-loop learning as well as the flexibility
to respond to the dynamics of uncertainty and disagreement (Patton, 2016b). This
question also asks evaluators to consider how the innovation and the evaluation
can co-evolve. Three elements are particularly relevant: the level of involvement
from various stakeholders (e.g., informing, consulting, involving, collaborating,
or empowering; International Association for Public Participation, n.d.); the di-
versity of perspectives and how to channel it in “active, reactive, interactive, and
adaptive” ways (Patton, 2016b, p. 257); and the complementarity of the evaluators’
and stakeholders’ skill sets.

FINDINGS: APPLYING THE FRAMEWORK TO TWO
ILLUSTRATIVE CASES

Overarching context
The cases we present are, in their own ways, innovations unfolding in complex,
dynamic environments. They originate from the RIPPLE program of research at
Queen’s University. RIPPLE’s mission is to explore research use and impact from multiple vantage points and perspectives, including research producers, intermediaries, and users. As evaluators in each case, we were involved as members of the innovation team to the extent possible, ensuring that aspects of the DE were co-created when practicable and that we could respond to issues as they emerged.

The specific cases were selected for two reasons. First, in view of Poth’s (2018) framing of a continuum of complexity in MM, these cases represent conditions of moderate to high complexity in that, for instance, establishing stakeholder capacities and implementing MM designs presented considerable challenges to our evaluative thinking and practice. Untangling the complexity of each case led us to a number of insights about how integration can contribute to advancing social innovations. Second, the cases provide a glimpse into how integration within a DE context can evolve under differing circumstances, such as scope (national vs. provincial), stakeholder composition, and leadership and funding structures. In what follows, we briefly describe each case context before applying our four-part framework to illustrate MM integration (why, what, when, and how) within and across the illustrative cases.

Case 1: Knowledge Network for Applied Education Research

The Ontario Knowledge Network for Applied Education Research (KNAER; www.knaer-recrae.ca) is an initiative to strengthen relationships between research producers and users in the K-12 education system to improve outcomes for students across four priority areas: mathematics, equity, well-being, and Indigenous knowledge. KNAER networks include over 100 partner organizations across the province of Ontario, including universities, school districts, community organizations, and policy partners. KNAER has a complex structure with many collaborative partners, including a Planning and Implementation Committee (PIC), Education Research and Evaluation Strategy Branch (ERESB), Coordinating Committee (CC), the Ontario Education Research Panel (OERP), the KNAER Secretariat, Knowledge Network (KN) leads, and Communities of Practice (CoPs) (Figure 2).

Case 2: Research Impact Canada

Research Impact Canada (RIC; http://researchimpact.ca/) is a network of 21 universities (20 in Canada plus the University of Brighton, England) that is sharing diverse methods for facilitating the societal impacts of research with local and global communities. The primary context for RIC is the Canadian research ecosystem, which functions as a mission-driven system for mobilizing research impact (i.e., impact is driven by researcher, institution, and funder goals). Within this context, RIC member universities possess considerable operational and geographic diversity. For example, while some members operate in large urban centers with resources spread across multiple individuals and units, others operate in rural contexts through the efforts of a single member. The diverse membership is united, however, by the innovative mission and activities of RIC to develop institutional capacity for mobilizing research evidence, which is unique among networked approaches to building such capacity.
Why was integration necessary?

While both cases operate with different scope—KNAER as a provincial initiative targeting the K-12 education sector and RIC as a national network working across sectors and disciplines—they share the core goal to increase research use for the benefit of Canadians. DE was particularly suited to exploring KNAER and its four KNs due to its focus on development and collaboration. A MM approach was also important, as we wanted to investigate the breadth and depth of network interactions through the respective integration of quantitative and qualitative data sources. Much of the emerging literature about knowledge mobilization suggests that co-creation and meaningful involvement of end users are important components to changing systems (Martin, 2010; Phipps et al., 2016). Similarly, DE revolves around the recognition that “participants should play a major role in goal-setting” (Patton, 1994, p. 313). We worked initially with the Ontario Ministry of Education to develop expectations around differential levels of inputs over the course of the DE as well as a logic model for the overall initiative. Our first phase entailed a scoping review to develop an evaluation framework, which was presented to the Secretariat, KN leads, and CoPs for feedback at a full-day event. Early on (February 2018), we introduced the DE to KNAER’s diverse stakeholder groups for a full day of collaborative work to explore the KNs’ innovation goals and the implementation of network activities.
across the province. Approaches included arts-based inquiry to explore and compare how each of the four KNs conceptualized their work and roles within the broader network.

The introduction of DE made it clear that capacity building would help clarify the goals of DE and how MM would be approached. As a follow-up to that meeting, we created a plain-language summary about DE, which detailed how it differed from other forms of evaluation and the multiple data sources we proposed to collect—this was a crucial step in generating buy-in for the evaluation across the networks and emphasizing that our team’s role was one of supporting the network rather than one of accountability. This was met with varying degrees of apprehension, depending on previous relationships. Where trust and working relationships had already been established, KN leads were on board; however, where no existing relationships existed, further meetings and negotiations were needed to begin to develop trust. Another assumption we made was that each KN did not already possess embedded evaluation expertise, but it turned out that two networks (equity and Indigenous knowledge) already had advanced evaluation capacity and plans within their strategic goals. Another challenge was that not all KNs had the same goals or wanted to approach network building or KNAER in the same way. As Patton (2016c) highlights, “disagreements, even intense conflict, among key stakeholders about what can and should be done can add to the turbulence and uncertainty” (p. 305). A key learning piece from this phase around goal setting and clarity was to build in room for contextual diversity to meet the needs of different partners, and to explicitly build in extra time for relationship building where new partnerships are emerging.

Similarly, in the case of RIC, the evaluation was positioned to support ongoing development, Patton’s (2011) first-listed purpose of DE. With this purpose crystallized, we resolved to ensure utility of the evaluation findings by first determining the methodological approach that would support RIC as the innovation. We began by examining the co-produced evaluation framework that had been developed in early 2018 by another evaluation team (see Bergen, 2019). The first element of the framework was a logic model linking RIC’s vision to its audiences, enabling conditions, common activities, and short- and long-term outcomes. The second element was a measurement overview linking the dimensions of the logic model to general data-collection priorities. To build familiarity with the evaluation framework and the RIC network, we reviewed previously collected quantitative event data (e.g., participant self-report data from earlier webinars) and supported the collection and analysis of qualitative data at the network’s annual meeting in September 2018. This initial work focused on building relational trust by engaging in existing evaluation efforts. An unanticipated benefit was that we observed initial evidence for the appropriateness of MM through the network’s readiness to engage with multiple types of data.

Following the annual meeting in September 2018, the lead author joined the RIC Evaluation Committee, ensuring that the evaluation was intertwined with the innovation. Over the ensuing four months, we reviewed findings from the
data mentioned above, contrasting emerging insights with recent scholarly and grey literature. Drawing from the evaluation efforts of the Michael Smith Foundation for Health Research\(^2\) (Penny Cooper & Associates, 2017), we determined two areas of focus to provide meaningful information about RIC’s activities: usefulness and use (MacGregor & Phipps, 2020). Usefulness addresses the quality of RIC’s activities in terms of members’ perceptions of their appropriateness, applicability, and practicality. Use addresses how members employ the information gained from those activities in their professional practice. With these foci (each constructed as a developmental question), our MM evaluation question asked, “For the network activities that RIC members find particularly useful for their professional practice (initial quantitative results), how do those activities contribute to their use of practices to mobilize research impact (explored in the qualitative phase)?”

**What was integrated?**

For the KNAER case, the DE methods were divided into six phases (Figure 3), many of which had points of integration, including a scoping review of research-practice partnerships to build a framework for evaluation and cross-network comparisons; a needs assessment of KN leads and key stakeholders (interviews) alongside an analysis of the data and metrics that networks were already collecting (document analysis of reports and implementation plans), integrated to build logic models for each network and refine the DE questions; co-creating DE plans and evaluation framework to compare the activities and impacts of KNs with different target audiences; surveys using social network methods to map the networks and nested communities within KNAER, integrated with further interviews with the KN leads to help understand the various patterns of interactions; and eight case studies at the CoP level to explore how teachers were using evidence-based KNAER resources to improve teaching and learning in classrooms across the province.

For the RIC case, integration was at the design level and followed an explanatory sequential approach (Creswell & Plano Clark, 2018), wherein the qualitative strand helped explain initial findings from the quantitative strand (Figure 4).

Initially, this meant two general phases: First, focusing on the usefulness construct, we co-created and deployed a survey to the Governance Committee (i.e., RIC members who oversee the network’s activities and strategic planning); second, focusing on the use construct, we used the survey results to identify key areas for developmental feedback to explore in the qualitative strand. The survey results also helped in building a sampling frame of RIC members (intended users) likely to provide information supporting adaptation of the network’s activities (intended use). Despite the apparent rigidity of design-level integration, we found success in the approach by explicitly planning for ongoing and fluid dialogue between the quantitative and qualitative strands, permitting the strands to evolve independently and jointly to improve alignment with the evaluation context, and keeping the intended use of findings foregrounded.
Figure 3. Procedural diagram for the developmental evaluation of illustrative case 1: Knowledge Network for Applied Education Research

*Note.* Timepoints of integration, when the quantitative and qualitative strands interfaced, are represented by arrows crossing from one strand into the other.
Figure 4. Procedural diagram for the developmental evaluation of illustrative case 2: Research Impact Canada

Note. Timepoints of integration, when the quantitative and qualitative strands interfaced, are represented by arrows crossing from one strand into the other.
As an example of keeping our MM design emergent and flexible, the “post annual meeting survey” depicted in Figure 4 was not included in our original planning. However, the confluence of several factors created an immediate need to know about RIC members' perceived outcomes from the annual meeting in addition to areas for improvement. These factors included the creation of RIC’s Management Team in September 2019, an inbound change in network leadership, and the growth of the network by three universities. Although data collection for the qualitative strand was also set to begin at that time (December 2019), we realized that generating insights from the semi-structured interviews would require more time than available to support imminent long-term planning. Rather than resist a change to our original planning, this situation created an opportunity for flexible integration. Structured interview questions originally considered for the interview protocol were adapted for the survey, creating space for more open-ended conversation in the interviews. With the survey and interviews implemented in parallel, we are now merging the quantitative and qualitative findings for analysis and comparison, identifying insights that would go otherwise unknown (e.g., RIC members were craving “ground rules” or best practices for engagement within and outside the annual meeting).

Integration is complicated, especially due to the range and diversity of integration objects that are possible. As such, cross-case analysis of integration objects for these evaluations draws on the dimensions from Fetters and Molina-Azorin (2017). Table 2 identifies the relevant dimensions and describes how each dimension was applied within each case.

A tension that arose in the design, data collection, and reporting was duplication of data that could occur across the levels of each network and the evaluation team. Hence, our goal was not to duplicate data collection but to orchestrate alignment across objectives and data sources through integration, providing a more holistic evaluation of how KNAER and RIC initiatives were positively influencing the evidence-use landscape. Another key piece of learning was to dovetail data collection at times that leverage existing reporting structures for each of the networks, in order not to be burdensome on the participating stakeholders, as we recognized the time constraints that evaluations can impose on participants.

When and how did integration occur?

In both cases, our approach focused on feedback mechanisms so that networks could learn and adapt from evaluative data sources and findings each year. Clarke and Fuller (2011) highlight that feedback mechanisms are essential for “corrective action, overlapping activities, and cyclical decision-making” (p. 91). Feedback loops were critical to gauging the KNAER’s progress longitudinally. Feedforward loops were also important to KN learning, offering a way to implement lessons learned going forward (Clarke & Fuller, 2011; Cooper et al., 2018; Reast et al., 2011; Vurro et al., 2011). De Montigny et al. (2017) highlight the fact that “rapid feedback loops can generate improved learning of unexpected opportunities and previously overlooked resources to produce quick wins and keep momentum
<table>
<thead>
<tr>
<th>Integration dimensions</th>
<th>Case 1 application: KNAER</th>
<th>Case 2 application: RIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theoretical approach</td>
<td>Collaborative approaches to evaluation (Shulha et al., 2016), DE (Patton, 2011), knowledge brokering (Cooper, 2014), and research impact theories (Kothari et al., 2011) were integrated.</td>
<td>The ongoing development (Patton, 2011) of RIC’s approach to developing institutional capacity for mobilizing research evidence.</td>
</tr>
<tr>
<td>Team</td>
<td>Evaluators specializing in collaborative approaches to evaluation, knowledge mobilization specialists, as well researchers who focus on K-12 education systems.</td>
<td>Knowledge mobilization scholars and practitioners, who collectively held expertise in DE, quantitative and qualitative methods, and the praxis of integration.</td>
</tr>
<tr>
<td>Literature review</td>
<td>Phase 1 included a scoping review of research-practice partnerships to create a framework to explore and compare the efforts of the four KNs. This was incorporated by introducing the framework to all stakeholders as a model for studying their collective efforts.</td>
<td>During the planning phase, an environmental scan of the extent literature identified the evaluation efforts of the Michael Smith Foundation for Health Research. As they note, and as we affirmed in further reading, “the literature supports the use of mixed methods, with a qualitative and/or narrative component providing rich understanding about how the [knowledge mobilization] intervention worked (or didn’t)” (Penny Cooper &amp; Associates, 2017, p. 19).</td>
</tr>
<tr>
<td>Rationale</td>
<td>Two elements of integration occurred in this dimension. The scoping review led to the use of particular metrics to assess the effectiveness of research-practice network. It also illustrated the value of using SNA to investigate the KNs.</td>
<td>We determined two areas of focus for providing meaningful information about RIC’s activities: usefulness and use. A MM design enabled us to expand upon the findings about usefulness when subsequently investigating use.</td>
</tr>
</tbody>
</table>
### Table 2. (Continued)

<table>
<thead>
<tr>
<th>Integration dimensions</th>
<th>Case 1 application: KNAER</th>
<th>Case 2 application: RIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data collection</td>
<td>Our intent of integration in Phases 1 and 2 was to expand our understanding and build case studies of each KN. In subsequent phases, using SNA in conjunction with interviews, the intent was cross-network comparisons to inform strategic plans for growth and teacher involvement.</td>
<td>We followed a two-phase explanatory sequential approach to collect data, enabling integration through building (Fetters &amp; Molina-Azorin, 2017). In Phase 1, we co-created and deployed a survey that focused the usefulness of networked learning activities. In Phase 2, we used the survey results to identify key areas for developmental feedback to explore in semi-structured interviews with key informants, focusing on the use construct. These overarching phases independently and jointly evolved.</td>
</tr>
<tr>
<td>Data analysis</td>
<td>Funding was cut prior to completing the SNA phase of the study, but the design included integration of data analysis (SNA survey and interviews with network participants) for the four KNs as well as their patterns of interaction with the Ministry and KNAER secretariat.</td>
<td>Our analytic approach can be characterized as intramethod analytics (Fetters &amp; Molina-Azorin, 2017), wherein the quantitative and qualitative data were analyzed within their respective strands before merging insights and informing subsequent data collection.</td>
</tr>
<tr>
<td>Data interpretation</td>
<td>One example of how integration provided an area of discordance was between the quantitative metric analysis and qualitative interviews for the needs assessment. While a major area identified in the interviews was impact on policies, the metric analysis uncovered that no data were being traced or collected in relation to this goal and impact area. As such, recommendations were made as to how KNs might trace their influence on the policy landscape.</td>
<td>Making meaning of the mixed findings has relied on the tiered involvement of various stakeholders within RIC, placing emphasis on ongoing feedback and co-creating outputs. Effectively channeling the complex involvement of participants meant that we regularly encountered a tension between providing rapid feedback and representing the diversity of perspectives.</td>
</tr>
</tbody>
</table>
Different feedback loops were established with different stakeholders at different times. Sometimes the evaluation team met only with the policymakers; at other times, we met only with the KN leads and managers. There was a lot of tension around the timing of feedback. Despite best efforts, we were challenged in analyzing data and producing recommendations at the speed that KNs needed. As such, we devised a plan to automate the data analyses through web-based dashboards. KNs would use an app or survey to input metrics that they were already reporting on (this function was meant to replace the existing reporting structure), and the dashboards would be built to provide not only network-specific data but also cross-KN analyses for the larger initiatives. A key takeaway from this work was the need to automate processes of data analyses in order to have the data influence decision making in a timely manner.

From the outset, RIC showed readiness for DE, particularly leadership buy-in, willingness to take risks, and flexibility toward involvement in the evaluation. We could thus pay full attention to ensuring co-creation and timely feedback rather than expend resources addressing common threats to the evaluation process (e.g., see Patton, 2016b). As in Case 1 Figure 3, we represent points of interface in Figure 4 (i.e., timepoints when the quantitative and qualitative strands were brought together) using arrows crossing the quantitative-qualitative divide. For RIC, there have been four points of interface, collectively forming the “temporal flow of activities” (Plano Clark, 2019, p. 109). However, representation of these points belies the dynamic, non-linear nature of how integration informed action. Giving primacy to ongoing feedback, we took to viewing these points as lodestars around which constellations of stakeholder conversations and integration actions were distributed. Each action or conversation has contributed in some small way to single-loop learning (i.e., informing action to address an identified issue), but it is the constellation of these that is reducing uncertainty and creating the potential for double-loop learning (i.e., “solutions to change the system itself”; Patton, 2011, p. 11). This viewpoint has supported the balanced perspective that learning processes “in emergent programs are likely to be subconscious, leading to haphazard and incidental learning” (Shea & Taylor, 2017, p. 87). By not focusing on double-loop learning as the sole goal for DE, informing action through integration has been gradual and interactive, though still in the direction of systems change.

With this strategy for generating learning through integration, we adopted a tiered approach to involving RIC’s various stakeholder groups. This approach ensured that the DE respected each stakeholder’s available time and the complementarity of their knowledge and skills for specific components of the evaluation. Every DE product was co-created, including the data-collection tools and reporting products, yet we realized early on that an everyone-at-once method for co-creation was more burdensome than useful. Each stakeholder group had a role to play in integration. As with Case 1, however, effectively channeling this complex involvement meant that we regularly encountered a tension between providing rapid feedback and representing the diversity of perspectives.
The central aim of this article was to consider the opportunities and challenges of MM in a DE context, with specific attention to the topic of integration. For those familiar with DE, it is perhaps unsurprising that our cases lend credence to Patton’s (2016a) contention that “the process and quality of engagement between the primary intended users (social innovators) and the developmental evaluators is as much the method of developmental evaluation as any particular design, methods, and data collection tools are” (p. 11). Simply put, in our experiences, MM does not make DE any easier, more appropriate for the particular social system, or likely to promote a utilization focus. As Mertens (2018) similarly notes, “not all evaluations take place in complex contexts, and not all evaluations require the use of mixed methods” (p. 8). However, when MM is appropriate, it can enable insights only achievable through integration. From the juxtaposition of our cases—why integration happened, what objects were integrated, and when and how integration was carried out—we conclude with three considerations for evaluation practice and future research.

1. **Balancing the need for integration with moving initiatives forward**

While MM provides unique opportunities to understand complex systems within DE, it also intensifies data collection. MM data collection can grind the progress of an initiative to a halt due to the high demands that are often placed on participants who have limited time within their full-time professional responsibilities for evaluative work. A potential solution is to spend time at the outset creating technological approaches that automate data-analysis processes: for instance, dashboards could pull from survey spreadsheets to continually analyze descriptive statistics over time. A survey process, especially where common metrics are needed over multiple events and timepoints, could also be built into an app that would feed directly into a dashboard, further simplifying data collection for participants in large, multi-stakeholder evaluations. These approaches are increasingly possible thanks to a range of technological and cloud-based programs, and even app development can be cost-effective when weighed against the cost of research assistants and teams doing that work.

The benefit to participants within the initiative being studied is that they can focus on keeping the work moving, rather than the evaluation itself becoming a major focus. For the evaluation team, a primary benefit is real-time data analysis that can be fed back almost instantaneously, addressing the issue that findings often lag behind the timeframes needed for them to be useful. Another opportunity to make MM integration feasible is using existing data sources (e.g., document analysis, text mining) to control the assorted costs of collecting and integrating qualitative and quantitative data. Data collection fatigue among participants can pose a serious issue, with evaluations sometimes collecting reams of data without focusing on how it will be used. It is critical to observe Patton’s (2016c) emphasis...
on being utilization-focused in order to limit and bound data collection. Moreover, being explicit with stakeholders about utility is essential; by being informed about why data is being collected and how it will inform learning, participants can see the value of engaging in the DE process.

2. Assuming stakeholders want co-creation

A key lesson learned from both evaluations was that stakeholders’ perspectives remain critical to the conversations of integration (between quantitative and qualitative data streams). It is often through engagement with stakeholders that integration insights are not only highlighted but also leveraged toward future action. Even so, the intensity of involvement in co-creation must be considered. In some cases, evaluation teams operate under an assumption that all participants want to be fully engaged in co-creation, from the research questions, to data and interpretation, to strategic planning meetings. For many groups and individuals in our experiences, that was simply not the case; they wanted to be involved at different levels depending on the phase and integration goals, and that was entirely acceptable.

In this respect, Boyer et al. (2018) developed a multi-level approach to stakeholder engagement for a clinical data research network. Their model works across two axes, where there is a greater number of stakeholders involved in short-term activities (surveys, interviews), moving toward higher levels of engagement with fewer stakeholders as time commitments increase. At the highest level of engagement, participants are embedded in the evaluation team and have advanced knowledge, expertise, leadership, and experiences to contribute. Such a multi-level approach to co-creation can be useful for mapping the involvement of various stakeholders over time for a DE, including their involvement in MM integration. We recommend building in time at the evaluation outset during which, for different phases and projects, people can explicitly choose their levels of involvement.

3. Cultivating integration literacy

Making effective use of MM approaches, and thereby integration, requires methodological expertise in both quantitative and qualitative traditions, as well as specialized knowledge about how to bridge their philosophical, methodological, and methods dimensions (Fetters & Molina-Azorin, 2017) in a way that is rigorous and timely. Scholars have repeatedly stressed the importance of research and evaluation teams being purposefully assembled with the requisite expertise to effectively employ MM approaches (e.g., Plano Clark et al., 2015). Without sufficient expertise, the increased complexity of MM designs can interrupt the heightened requirements of DE, such as ensuring rapid feedback to promote use. For example, an under-experienced evaluator may struggle in determining what objects to integrate in an evaluation. We thus contend that an important direction for future evaluation research and practice will be to explore how evaluators considering or employing MM approaches can build their integration literacy, which we take to mean an ability to determine why integration is appropriate, what objects of
the evaluation should be integrated, and when and how integration will occur. In the end, both cases involved stakeholders who were willing to take risks and participate within innovative DEs, as well as MM integration. And in each, it was the willingness and expertise of the stakeholders themselves that contributed to successful evaluations that pushed cyclical rounds of network improvement for the good of the end-users: Canadian citizens.

AUTHOR INFORMATION

Stephen MacGregor is a SSHRC Postdoctoral Fellow at the Ontario Institute for Studies in Education, University of Toronto. His research focuses on how the lessons learned from multi-stakeholder networks can be mobilized to achieve impacts upon wider society. ORCID: https://orcid.org/0000-0001-7466-5695

Amanda Cooper is the Associate Dean of Research and Strategic Initiatives at the Faculty of Education, Queen’s University, and the founder of RIPPLE (Research Informing Policy, Practice and Leadership in Education), a program of research, training, and knowledge mobilization aimed at learning more about how knowledge brokering can increase research use and impact in education (www.ripplenetwork.ca). ORCID: https://orcid.org/0000-0002-6880-9047

NOTES

1 Throughout this article, we use “mixed methods” as a general term, yet we recognize that “mixed methods research” and “mixed methods evaluation” are distinguished by the purpose and political context of an evaluation (Mertens, 2018).

2 A health research funding agency in Canada with a focus similar to RIC on improving the societal impacts of research (https://www.msfhr.org/about).

3 The KNAER DE was cancelled unexpectedly when a new government came into power, so some of the design elements we discuss in the paper were not implemented, as the evaluation was cut short despite those phases being planned.

4 Programs considered included Airtable (www.airtable.com) and Tableau (https://www.tableau.com).

REFERENCES


doi: 10.3138/cjpe.71488


