Evaluating a complex research capacity-building intervention: Reflections on an evaluation of the African Institutions Initiative

Sonja Marjanovic[1], Gavin Cochrane[1], Enora Robin[2], Nelson Sewankambo[3], Alex Ezeh[4], Moffat Nyirenda[5], Bassirou Bonfoh[6], Mark Rweyemamu[7] and Joanna Chataway[1,8]

Abstract

There is increasing policy demand for real-time evaluations of research and capacity-building programmes reflecting a recognition of the management, governance and impact gains that can result. However, the evidence base on how to successfully implement real-time evaluations of complex interventions in international development efforts is scarce. There is therefore a need for reflective work that considers methodologies in context. This article shares learning from the experience of conducting a participatory, real-time, , 'theory driven' evaluation of the African Institutions Initiative, a Wellcome Trust funded programme that aimed to build sustainable health research capacity in Africa at institutional and network levels, across seven research consortia. We reflect on the key challenges experienced and ways of managing them, highlight opportunities and critical success factors associated with this evaluation approach, compared with alternative evaluation approaches.

Keywords

[1] RAND Europe, UK
[2] University College London, UK
[3] Makerere University, Uganda
[4] African Population and Health Research Centre, Kenya
[5] Malawi Liverpool Wellcome Trust Clinical Research Programme, Malawi
[6] Centre Suisse de Recherches Scientifiques en Côte d'Ivoire, Côte d'Ivoire
[7] Sokoine University of Agriculture, Tanzania
[8] Open University, UK

Corresponding author: Dr Sonja Marjanovic, Research Leader, RAND Europe, UK. [email: smarjano@rand.org] [author: please provide postal address]

Introduction

Evaluating complex interventions

Effective evaluation approaches are influenced by both the specific purpose of an evaluation and by the complexity of the intervention being evaluated. Literature distinguishes between simple, complicated and complex interventions (e.g. see Glouberman and Zimmerman, 2002; Campbell et al., 2007; Rogers, 2008; Bamberger et al., 2016, see also Supplemental Data 1 [Sage: please add the necessary info/link]), based on a range of interrelated criteria including: the environment in which the intervention is being developed and deployed; up-front knowledge about success or failure factors; the complexity of the intervention in terms of different constituent elements; and the feasibility of pre-specifying all possible outcomes (e.g. Ling, 2012).

The more complex the intervention, the more difficult will be the evaluation (e.g. Gadsby, 2011). Complex interventions are generally not conducive to randomised controlled trial (RCT) methods. This is in part related to their intrinsic diversity and to difficulties in establishing clear counterfactuals or comparators. The problem(s) that a complex intervention targets tend themselves to be multidimensional and difficult to fully specify. Lack of up-front knowledge or agreement on how change processes will unfold and on the early and mid-term changes which need to happen for the long-term vision of a complex program to be achieved (Weiss, 1995) further accentuate these evaluations.

In international development settings, in which this article is situated, the need to provide evidence on 'what works' when evaluating complex interventions has been met by a growing recourse to various 'theory' based evaluation approaches.. These include 'theories of change' (Weiss, 1995; Connell and Kubisch, 1998), and 'intervention logics; 'Realist' evaluations that emphasise the importance of context and mechanism – asking not only what works, but for whom and under what circumstances (Pawson and Tilley, 1997); and 'contribution analysis' which focuses on how a programme contributes to rather than 'causes' outcomes (Mayne 2012). Such theory driven evaluation approaches that follow through the links in the chain of intervention, implementation through to outcomes and impacts lend themselves to real-time evaluation during a programme's life rather than waiting to the end. This can facilitate timely, formative feedback needed for the

adaptation of an intervention, allowing the evaluation to be sensitive to changes in direction and unintended consequences.

There is limited evidence on how to successfully implement real-time evaluations that are theory informed , despite their increasing use when evaluating complex interventions. Some studies describing such evaluations exist, but few discuss the merits, limitations and ways to mitigate challenges associated with such evaluation approaches. In addition, given the policy-oriented nature of evaluating complex interventions, which often occurs under financial, time, political and data constraints, there is a need to reflect on the practicality and feasibility of these approaches (Bamberger et al., 2016).

This article aims to enrich methodological and practice knowledge by learning from our experience of conducting a real-time, theory -driven evaluation of a complex health research capacity-building intervention: the Wellcome Trust-funded African Institutions Initiative (AII).[1] The initiative funded a network of seven consortia composed of research organisations and universities. It aimed to build sustainable research capacity in Africa at institutional and network levels through African ownership and control of capacity-building efforts. This article focusses on the evaluation process itself, rather than on the findings about enablers and challenges of research-capacity-building more widely.[2] The authors of this article represent both members of the independent evaluation team, commissioned by the Wellcome Trust, and Directors of the African research consortia participating in the programme. As discussed later in the article, the independent evaluators and the AII partners co-produced the evaluation framework, and worked together on its refinement, implementation and interpretation of evaluation data over time.

In what follows, we first introduce some of the challenges associated with evaluating complex research capacity-building interventions and then introduce the AII as such an intervention, describing the evaluation approach and associated methods that were used. Drawing on the experiences of both the independent evaluators and of the AII partners in the evaluation, we then discuss the key evaluation-related challenges encountered, and consider how they can be managed. We conclude with a reflection on the opportunities and critical success factors associated with this evaluation approach, but also discuss some alternative options for initiatives that share features of AII complexity.

Evaluating complex research capacity-building efforts

Policymakers and funders are showing increased demand for evaluation of research and capacity-building programmes (Gadsby, 2011). This reflects a broader recognition of the management, governance and impact gains of evaluation. It is also motivated by the need to demonstrate accountability for investments, showcase successes (e.g. for advocacy), enable learning about critical success factors, and to help inform future strategy (cf Marjanovic et al., 2009). Sharing experiences from prior evaluations can also help efforts to coordinate evaluation activities between funders in the international development field. For example, the ESSENCE[3] on Health Research initiative aims to develop common frameworks for planning, monitoring and evaluation of capacity-building activities (Boyd et al., 2013).

However, research capacity-building interventions can be particularly difficult to evaluate, especially in a participatory manner. Capacity-building is often seen as a subjective attribute and highly context specific. Challenges to evaluating research capacity-building are all the greater in international development contexts, where researchers able to engage with evaluations are often highly overcommitted (e.g. see Trostle, 1992). Balancing the needs for evaluating short-term impact and progress (as part of an ongoing learning process), with the establishment and implementation of tools and methods needed to assess longer-term achievements and sustainability is also not straightforward (Bates et al., 2014). For example, sustaining appropriate levels of stakeholder engagement to ensure the feasibility of an evaluation and consistent findings and inferences is a resource-, time- and relationship-intensive process.

Thus, while there is a growing body of influential literature on research capacity-building (e.g. Horton and Mackay, 2003; Whitworth et al., 2008; Ezeh et al., 2010; Marjanovic et al., 2012; Vasquez et al., 2013; Chu et al., 2014), there is comparatively little comprehensive evidence on how to effectively *evaluate* research capacity-building. Many evaluations have been conducted and provide useful learning on the process and impacts of the research capacity-building endeavour (e.g. SIDA, 2005; Bates et al., 2006; Jones et al., 2007; NORAD, 2009; Minja et al., 2011) but there is much less reflection on the challenges of the evaluation process itself, and especially on how to manage them in development contexts.

Existing literature on the evaluation of research capacity-building interventions tends to focus on design (e.g. identifying what to evaluate and which indicators to use, Cooke, 2005) or on the evaluation process (Neilson and Lusthaus, 2007). Less attention is paid to the nuances of the evaluation context, which influence whether a complex intervention and its evaluation are successful or not. An exception is literature reflecting on a very specific element of the evaluation process i.e. the theme of participation. Some international development-focussed monitoring and evaluation manuals identify common pitfalls and challenges (e.g. Stern et al., 2012) in capacity-building and development evaluations, such as those related to the attribution of benefits or ownership of the evaluation. However, insights on how to manage or resolve the challenges are limited, patchy and often difficult to use. Bates et al. (2011) and Cole et al. (2014) examined aspects of capacity-building evaluation and concluded that evaluation frameworks and associated indicators do not look into the inter-relatedness of different activities and resulting outputs and outcomes. This strengthens the argument for theory-based approaches which focus on the sequence of events connecting processes to outputs and outcomes.

In what follows, we briefly introduce the African Institutions Initiative as a complex intervention, and then describe how it was evaluated, to lay the foundation for subsequent discussion of the evaluation method-related lessons learnt.

Evaluating the African Institutions Initiative as a complex intervention: design and methods

The African Institutions Initiative as a complex intervention

The AII was an innovative and large-scale example of a growing number of networked research capacity-building initiatives that have emerged in response to the need to expand research capacity Africa. Established in 2009 for an initial period of five years, the initiative funded seven interdisciplinary health research capacity-building consortia incorporating initially 51 institutions in 18 African countries, and 17 partners across Europe, the United States, Australia and Malaysia

(See Figure 1). The initiative promoted individual and institutional capacity-building through a range of activities:

- individual training (Postdoc, PhD and MSc scholarships) the development of improved postgraduate curricula and the introduction of new research positions and training programme structures, as vehicles for institutional change in established practices;

- strengthening research-career prospects and the profile of research in African universities (e.g. advocacy, professional development training, small grants);

- improving research governance, management and administration capacity (e.g. research management training, recruiting support posts, introducing new governance and management structures, practices and policies for supervision, student selection, financial reporting, evaluation);

- improving the physical infrastructure (e.g. ICT, new lab equipment); and

- strengthening collaborative networks between individuals and organisations across geographies (e.g. conferences, student and staff exchanges, joint supervision).

[Figure 1]

The AII can be considered a complex intervention for a number of interlinked reasons. First, there were many uncertainties in the intervention context, and in relation to this, limited and fragmented evidence on success factors. One example of such uncertainty related to political turbulence and socioeconomic instability in some participating countries, which had impacts on the programme's evolution (e.g. Bonfoh et al., 2011). Second, there were many interdependent components which needed to function together for the initiative to deliver on its goals. These include individual empowerment, training programme strengthening, research governance and management capacity-building, infrastructure improvement, network connectedness and support for institutions with both weaker and stronger initial capacities. Third, the intervention and its context were also highly interdependent. For example, in the evaluation of the AII, this was evident in how efforts to institutionalise postdoctoral positions and to advocate merit-based promotion pathways within universities unfolded. These efforts to establish new or changed systems were embedded within

pre-existing university career structures, working practices and national science policies and were influenced by them. Finally, the ability to specify the full range of potential outcomes at the outset was limited as there was a high propensity for adaptation and change. For example, adaptations in the degree of emphasis some consortia placed on capacity-building at different stages of research career pathways (Postdoctoral, PhD or Masters) changed over time, in response to local recruitment circumstances.

The overarching conceptual framework used in the evaluation of the African Institutions Initiative

As introduced earlier in this article, the evaluation of the AII was rooted in a theory- driven, real-time approach. In what follows, we provide an overview of how the approach was implemented in practice in order to provide sufficient background information and context for a subsequent discussion of the opportunities and challenges associated with implementing such an evaluation.

Each consortium in the AII, as well as the initiative overall, developed theories-of-change and associated logic models with a combination of common and unique features. Common features included: efforts to build capacity at individual and institutional levels through local leadership and a networked approach; an emphasis on scientific skills, professional practices in research governance, management and administration; and infrastructure strengthening. Some of the unique features included: differences in disciplinary and field focus; the mix of capacity-building interventions being implemented; and the levels of funding awarded.

Each consortium's specific objectives and their implementation approaches, as well as the initiative's overall objectives and the learning aims of the evaluation, influenced the establishment of a framework to guide the evaluation process (Figure 2). This framework served as a means for learning about the initiative as a whole, by bringing together the experiences of each consortium. It is important to emphasise that the framework was developed as a high-level evaluation tool to bring out commonalities and that each consortium was also evaluated against their own unique interventions. One of the key considerations in developing the framework was the need to reflect that the consortia shared some common features but were also intrinsically diverse.

[FIGURE 2]

The framework was established at the onset of the evaluation through collaboration between the independent evaluation team and AII consortia. The independent evaluators analysed background documentation and consulted with consortia members, on the strategic approaches of each consortium in the initiative. The framework consisted of three overarching common '*categories of effort'* covering both institutional and individual capacity-building goals across the consortia. These categories of effort reflect the core areas of capacity-building that were being targeted during the programme's life-span and revolve around (see Figure 2):

- *Capacity-building in scientific skills and research training.* Training and empowering individuals to conduct research; strengthening institutional receptiveness such as career development prospects at universities);

- *Improving research governance, management and administration capacity.* Training individuals in grant-writing, financial management, ethics, project management, supervision, publication writing; and implementing better knowledge management systems;

- *Improving ICT and physical infrastructure.* Investing in research infrastructures based on a critical assessment of institutions' specific needs; sharing of available infrastructure within institutions and between projects.

In addition to these three categories of effort, networking, linkage and exchange were important cross-cutting elements of the initiative and the evaluation team also assessed the way in which the programme's networked approach functioned. The evaluation used these categories in the conceptual framework: to facilitate the evaluation and learning process; to ensure a requisite degree of consistency in approach for cross-consortia learning (e.g. about enablers and barriers to delivery and successful capacity-building approaches); and to enable the unique activities and operational contexts that characterise the theories-of-change of individual consortia to be captured in a systematic way. For each category of effort, the independent evaluators and AII evaluation partners considered whether a consortium was delivering on commitments (feasibility and efficiency of the approach), and whether the activities selected to pursue consortium objectives were leading to desired outputs, outcomes and impacts (effectiveness of the approach,

as well as utility and sustainability). Specifying the intervention logic for each consortium helped in developing a dashboard of quantitative and qualitative performance indicators in key categories of capacity-building activity. Process indicators and measures clarified whether things went according to plan, and enabled the exploration of underlying reasons. Output and outcome indicators helped us identify what was being produced and to what extent the activities contributed to the longer-term goals and impacts. Together, the diversity of indicators and the accompanying narratives helped us verify the underlying assumptions in the programme.

The independent evaluators and the evaluation partners in the consortia worked together to try make the framework adaptable to addressing the need for learning from comparable elements and for accounting for unique aspects of different consortia. The challenges experienced, some associated with different stakeholders needing different types of information and learning from the evaluation, are discussed further below. For an overview of evaluation questions and criteria, please see Supplemental Data 2. [Sage: please add the link]

Implementing the theory of change approach in practice

The independent evaluators and evaluation partners in the consortia sought flexibility in the evaluation design in order to balance the multiple purposes of an evaluation (accountability, learning, advocacy, informing strategy); the mix of common and unique features across consortia; and the multidimensional nature of the AII (i.e. the focus on individuals, institutions, networks). The evaluators also opted for a participatory approach in which consortia were engaged with the design and implementation of the evaluation. Although the appropriate degree of participation (and how to balance its merits against a desire not to interfere unduly with programme implementation or overburden participants) remains a challenge for many real-time evaluations, this evaluation evolved from an initial phase of intensive consultation to more regular and engaged active participation by the community of programme stakeholders.

The timeline of the evaluation and core associated components are outlined in Figure 3.

[FIGURE 3]

Component 1: Framework co-development, baseline assessment and milestone setting. The first objective was to establish rapport and good working relationships between the independent

9

evaluator and the consortia, develop a shared understanding of consortia goals and work plans, and co-produce the evaluation framework outlined above. To this end, the independent evaluators facilitated inception workshops with each consortium and their key stakeholders. Following background research and remote conversations with consortia leaders and the Wellcome Trust, initial framework development workshops were conducted on-the-ground, and attracted participants from each consortium spanning Directors, researchers, administrators and local evaluation managers. Each workshop took place in a partner country in Africa and lasted between one and two-and-a-half days, depending on how much time consortia could free-up around previously scheduled annual meetings. These workshops helped specify the intervention logic in a participatory way and created a platform for discussing and agreeing on an initial set of evaluation indicators. Each consortium also worked with the independent evaluator to ensure that evaluation protocols (e.g. interview questions, indicators used) were fit for purpose and could be adapted to reflect changes in the programme's theory-of-change and priorities over time. Refinement and finalisation of the evaluation framework also involved desk research (document review) and remote follow-up consultations between the independent evaluators and representatives of each consortium.

The evaluation also assessed consortia 'starting points' (i.e. baseline research capacity), in order to be in a position to accurately interpret evaluation evidence and contextualise progress over time. Baseline capacity assessments were coordinated by lead institutions in each consortium but involved all partner institutions (54 institutions – university research groups, faculties and departments, as well as research institutes – across 18 countries in total). Some initial information on baseline capacities was gathered during the inception workshops. A baseline survey was developed soon after, so that the questions could reflect consortium's theories-of-change and related indicators.[4] The baseline surveys gathered both qualitative and quantitative information, with questions balancing concern for methodological rigour with a pragmatic acknowledgment of the working realities of each consortium's data environment. For example, different levels of capacity amongst AII organisations influenced both what type of data was available, and how up-to-date it was. Although the initiative baseline date reflected the award of the AII contract, in practice different consortia moved to implement the programme at different times and with varying

delays. Consortia's own assessments of relevant and robust data had a strong role in the baseline assessment process and the evaluation team needed to make considered trade-offs between methodological rigour and relevance. Although the baseline data was not without its limitations, the evaluators believe that it provided a solid foundation against which to assess progress aligned to the operational realities of the evaluation context. The initial consortium workshops also provided opportunities for capacity-building in evaluation skills, although it is important to highlight that initial evaluation capacities varied across consortia. This was done through a combination of presentations on methods and learning-by-doing. Consortia also set annual milestones for key aspects of activity so that progress against plans could be reflected on and learnt from, and were invited to revisit these and the wider logic model each year. When evaluating complex programmes there is a need to acknowledge emergence, adaptation and change. Therefore, the evaluators did not consider milestones as set in stone and made explicit that not meeting milestones does not indicate failure. The evaluation team also facilitated sessions on risk identification and management, and strengths, weaknesses, opportunities and threats analysis.

Component 2: Ongoing evaluation and interim reporting. Each year, the independent evaluation team led an annual reporting process. Local evaluation leads within consortia managed and coordinated the process of data collection from partner institutions within consortia, provided guidance on feasible units of analysis and indicators, and participated in the dissemination and use of evaluation evidence in the region (circulated reports and presentations, in some instances helped organise workshops and evaluation meetings). Consortia collected data on agreed indicators annually, and produced quarterly summary reports on progress (shifting to biannual later in the initiative's life). The independent evaluator critically reviewed the indicator data for clarity, consistency and backing evidence. The consortia reviewed the interpretations of data made by the independent evaluator at annual intervals. The draft reports from the independent evaluators were shared with consortia Directors and evaluation leads for additional inputs, reviews and scrutiny on inferences, as well as consultation with the wider network of each consortium.

Interim findings and the evaluation process and protocols were discussed with consortia representatives and an initiative Advisory Board at annual initiative-wide meetings, learning was

shared in individual consortia sessions and in plenary, and amendments made as appropriate for the following year. These annual discussions and quarterly communications between the independent evaluators and consortia evaluation leads also led to adaptations in both the qualitative interview protocols over time and in the prioritisation of quantitative indicators based on relevance and feasibility.

Following discussions on implementation challenges, evaluation leads within each consortium were awarded a small funding supplement from the Wellcome Trust to facilitate enhanced engagement with the evaluation.

Component 3: Networking and dissemination for learning and exchange. Ensuring effective and timely learning is vital for strengthening research capacity-building efforts (Maselli et al., 2006). Annual initiative-wide meetings which attracted consortia representatives, regional stakeholders from academic circles as well as some policymakers and advisors were key to the exchange of insights and enabled formative feedback. Ongoing remote communications (email, telephone), interim publishing[5] and presentations at two international conferences[6] also facilitated linkage and exchange. When invited and when partial costs were covered by a consortium,[7] independent evaluators delivered supplementary workshops.

Component 4: Final assessment and reporting. A final report assessed progress against individual consortia objectives and for the overall initiative over its lifespan.[8] To do so, it drew learning from the annual assessments and validation at a final initiative-wide meeting. It also drew out thematic learning and provided recommendations for future efforts.

Capturing the learning: key challenges, options for their management and opportunities for future evaluations

Some of the challenges that were experienced as part of the evaluation were inevitable, given the scale and nature of the evaluation and the context in which it was taking place. With the benefit of hindsight, others could have been at least partially mitigated. Here, we discuss key evaluation challenges, based on the perspectives of both the independent evaluators and the consortia evaluation partners. At the end of each of section, we draw on learning from this evaluation, as

well as the evaluator's experience with other complex interventions, to present some options to consider in future evaluation efforts.

To identify key methodological challenges, we drew on various sources. Challenges and ways of addressing them were discussed between evaluators and each consortium during individual reflection and learning sessions at initiative-wide annual meetings. They were also discussed in plenary sessions including all consortia, the independent evaluators, the funder and Advisors. This facilitated similarities and differences between individual consortia experiences to surface and shared learning about mitigation and management. These events enabled initial classification of emerging insights into key themes representing categories of challenges and associated management insights. As part of the process of co-producing this article, the themes and management insights were verified through online communications between the independent evaluators and contributing co-authors from across the consortia. Given that this is not an analysis of insights pertaining to capacity-building processes, outputs and outcomes but rather a reflection on the methodological approach of the evaluation, we adopted Patton's (2001) 'conceptual use of findings' to deepen our understanding and increase shared learning based on the experience of the evaluation from multiple perspectives. The triangulation of insights across these sources of evidence identified the following categories of challenges, related to (i) a participatory approach; (ii) culture and history; (iii) operational issues; and (iv) managing multiple expectations and interests.

Challenges related to a participatory approach

Participatory approaches to real-time evaluation can help improve the appropriateness, acceptability and integration of evaluation evidence into programme activities. However, questions about how much participation is needed, the best methods for enabling constructive engagement and what pre-existing institutional arrangements influence participation need to be considered (see Chouinard & Cousins, 2012; Cousins & Chouinard 2013)

Striking the optimal balance between desirable participation levels and what is feasible and efficient given the operational realities of an evaluation (e.g. resource availability, timelines) is not straightforward. Substituting on–the-ground interactions with remote interactions, such as phone and email, is not a simple *quid pro quo*. Online participation creates opportunities for cost-savings

and a focus on prioritised data gathering. However, it can also limit an external evaluator's understanding of the local context and of intervention complexity. This can in turn result in additional time-demands associated with remote relationship-building and clarifications. However, simply building in very large amounts of on-the-ground time does not guarantee that such problems are mitigated. One consortium Director suggested a balance may have been struck by attending one annual meeting of each consortium, in addition to the attended initiative-wide meetings. Time has to be carefully planned and managed to ensure that all relevant voices are represented in communications. For example, more opportunities for face-to-face discussions between evaluators and partner institutions in consortia (as opposed to predominantly lead institution representatives) could have helped encourage greater understanding and engagement with evaluation evidence across the programme. Such opportunities were in part impeded by significant resource constraints for fieldwork and time-related constraints. This highlights another feature of participatory, real-time evaluations. They are expensive. Drawing on the experience of trying to conduct this evaluation for far less, the evaluators estimate that such an evaluation should represent approximately 8-10% of a programme's budget.

Different attempts were made to mitigate the tension between desirable and feasible levels of engagement in the AII evaluation. The independent evaluators supported the funder and consortia to establish specific roles for local evaluation oversight and coordination. Remote communications between the independent evaluator with the lead institution's evaluation officer and Director took place at regular intervals (to minimise burden but facilitate internal planning) and annual initiative meetings offered an opportunity for face-to-face interaction. The evaluators encouraged the lead institutions in each consortium to share emerging insights and learning widely across their networks. Responses to these efforts over time suggest that this was helpful.

Participatory, real-time evaluation also face questions about the degree of independence an evaluator assumes, and how decisions on this are made. Thus, during the course of this evaluation, a consortium member asked the evaluation team for advice on how to respond to underperforming students. Providing a direct answer would have been outside the independent evaluator's remit, but sharing learning about how other initiatives had addressed similar issues

and asking questions about performance review processes in their institutions was helpful and considered appropriate.

Our experience points to a number of actions which are worth considering as part of efforts to mitigate and manage the challenges associated with a participatory, real-time approach:

- *Investing sufficient time upfront to build a shared understanding of the evaluation approach and of its participatory nature, and to ensure the feasibility of implementation*. Related to this, the experience of the AII evaluation highlights the importance of discussing time and resource demands at an early stage in the process, and of engaging the funder in ensuring that those being evaluated are supported by appropriate financial and staff resources. Phasing requests for participation and providing sufficient upfront notice is important to avoid overwhelming demands.

- *Consulting with stakeholders to prioritise where participation can bring the most value across evaluation framework design, implementation and inference-making stages and how time-demands can be most effectively managed.* This includes being transparent about the trade-offs between breadth and depth, which may be required to ensure feasible levels of participation. It also points to the importance of discussing the balance between fieldwork and remote communications with the evaluation funder and with programme stakeholders at the onset.

- *Seeking participation from all relevant stakeholders, not only those being evaluated but also wider stakeholders who can influence the impact of capacity-building activities (e.g. University Vice-Chancellors, Ministries)*. In addition, it is advisable to triangulate sources of evidence. In the evaluation of the AII, we aimed to achieve this through testing inferences and consulting on the evidence with senior consortia leadership as well as evaluation officers and consortia project managers.

- *Being as adaptive and responsive as possible to the emerging needs, changing priorities and modifications in the theories-of-change of those being evaluated.* This is integral to evaluations of complex interventions given the high interdependence between the intervention and its context and the propensity for change in the intervention itself. In this evaluation, it was manifested through adaptations in evaluation protocols and some

indicators over time to ensure that appropriate questions were being prioritised. Such adaptability can help ensure that the evaluation process is feasible, relevant and strikes an appropriate balance between breadth and depth. However, it also implies for the need to negotiate the scope of evaluation activity at the outset with the funder and revisiting it regularly in light of emerging questions, priorities and lessons learned.

The barriers to addressing many of the challenges outlined above are associated with financial and time demands. This accentuates the importance of early communications, joint planning, and transparency in trade-offs between different levels of participation.

Cultural and history-related challenges

Evaluators, funders and those being evaluated often operate within and across very different cultures. Such cross-cultural diversity was evident in the AII on multiple fronts: between consortia partners from different African countries, between partners in Africa and Europe, between partners from different European countries, and between the independent Europe-based evaluator and consortia. The diversity can be a productive provided there is awareness and sensitivity to cultural and historical differences.

The evaluator has to balance many roles: an independent and impartial actor, but also a trusted partner, and on occasion a conduit or intermediary for managing and reconciling expectations, beliefs and interests between different stakeholders. These challenges were accentuated by the novelty and experimental nature of the African-led and cross-cultural capacity-building approach. For example, during the initial stage of the project some consortia felt that the demands of the evaluation were not in line with their priorities in often very difficult research, recruitment and political environments. This was accompanied by the risk of some consortia perceiving the evaluation as micro-management by a European institution raising historical sensitivities of neo-colonialism. This underscores the importance of developing and adapting methodological practices to the culture, context and values within which the initiative was operating (Chouinard, 2014).

Our experience of conducting an evaluation of this scale and nature, and managing challenges associated with working across different and diverse cultures, suggests that the following can be helpful:

- *Ensuring that evaluators are familiar with local evaluation contexts, cultures and associated sensitivities, and tailoring tone and discourse accordingly.* Ideally, a local evaluation partner would be best placed for this. Our experience suggests that external evaluators should share experience of previous work within different country contexts, with those being evaluated and local collaborators. Site visits are also important. In the AII evaluation, time on-the-ground was important for the external evaluators, consortia leads and evaluation officers in each consortium to get to know each other, and to gradually build shared understandings.

- *Investing substantially in early relationship-building – both to discuss the potential benefits of evaluation and understanding the motivations and constraints of those involved.* Some of those constraints may be associated with differences in working cultures and contexts including the different responsibilities a researcher may need to assume as part of their job, in the resource-constrained recruitment environments of many developing countries.

- *Investing time and effort in responding to criticism.* Although this appears obvious, regular engagement with criticism (for example, on issues related to time pressures and prioritisation), and clarity and openness in communication on *how* issues were being addressed were particularly important to establish relationships of trust.

Operational challenges to timely evidence

The most widespread operational challenge experienced in the AII evaluation related to the timely production of evaluation evidence. These occurred for a variety of reasons, ranging from competing time demands to political environments in the countries where consortia were located. Consortia noted that they often faced trade-offs when balancing the delivery of programme activities with meeting evaluation requirements. This was accentuated because evaluation data was needed from institutions across each consortium's network, rather than from a single lead institution. This required significant coordination efforts by consortia evaluation officers as well as a need to build a shared understanding of the evaluation across geographically dispersed partner institutions some of which had not collaborated with each other before. Greater clarity on the

evaluation requirements at the outset of the programme may have reduced the time needed for such communication, by potentially enabling more appropriate planning and resourcing.

Regular communication between the independent evaluators and consortia evaluation officers, flexibility in timelines for producing evaluation data, and an openness to iteration and clarifications helped tackle operational challenges and enabled improvements in data quality over time. Committed consortia leadership and evaluation officers were key drivers of improvements. Supplementary evaluation funding, provided by the Wellcome Trust to consortia, also helped responsiveness to operational challenges.

In summary, we offer some management devices that might potentially assist evaluations facing similar challenges:

- *Evaluators should work with the funder and those being evaluated at the outset of the initiative, to help ensure that there are:*

  o Adequate budgets for real-time evaluations of complex interventions to facilitate engagement: real-time participatory approaches are much more resource-intensive than ex-post evaluations (and especially of less complex interventions);

  o clear lines of commitment, responsibility and accountability for delivering on evaluation requirements, with staff having requisite financial support;

  o agreements on reporting timelines and how the reporting process will be managed. It is important to build in some flexibility in the timelines and to maintain open lines of communications around delivery.

- *Operational realities and their impacts on an evaluation need to be revisited regularly with all stakeholders involved, and a collective way forward negotiated.* In the AII evaluation, communication with the funder on the feasibility of collecting certain types of evidence - given variation in institutional record-keeping and time-demands on consortia - supported agreement on evaluation priorities and clarity on why some types of evidence could not be obtained.

Challenges related to managing multiple expectations and interests

The final set of challenges that were experienced related to stakeholder management, specifically to managing various interests and harmonising priorities. To the best of our knowledge, a participatory, real-time evaluation of a complex intervention of the AII's scale - involving 54 African and 20 non-African institutions, as well as other external stakeholders - has not been attempted before Funders, consortia, advisors and evaluators had some shared but also diverse expectations of the evaluation and the intervention. Expectations evolved over time and were not always clearly articulated.

Stakeholder learning about evolving priorities and interests took place throughout the evaluation and at multiple levels. This was important for adaptability, but led to trade-offs in terms of the stability and embedding new ptactices. Negotiation was required in light of a mix of complementary and competing interests of different groups, and limited evaluation resources. For example, in the AII evaluation one stakeholder group was particularly interested in gender whereas this was not a priority for most of the other stakeholders. This emphasises that it is important to make explicit that there are limits to what can realistically be expected of an evaluation, whatever methods are used, especially given the complexity and ambition of many interventions in developing countries.

Drawing on the experience of the AII, the following should be considered when evaluating interventions with similar levels of stakeholder complexity:

* *Discuss expectations with all parties as part of the commissioning process and contract award, and revisit regularly*. This includes specifying what is and is not included in evaluation scope as well as the expectations of all parties of their participation in the evaluation. Ensuring continued, regular engagement and reflection on expectations is important to sustain stakeholder buy-in and a shared appreciation of what is deliverable. It is also essential for partnership relations and for mitigating the risks of donor-recipient hierarchies.

- *Facilitate the funder in their own process of articulating and clarifying expectations, by asking specific questions*. For example, this could include questions about priorities, ways of managing potential risks and the trade-offs they might imply for the evaluation.

- *Be explicit that concrete recommendations will be provided for some issues, but that the evaluation is also likely to raise a number of questions for which there may not be a straightforward answer*. This is especially true given how little is known about success factors for complex interventions. For example, there is little evidence on the time needed to build institutional capacity for research, given the lack of a counterfactual and the inherent heterogeneity of the intervention.

- *Establish an Advisory Board early on*. The Board should include evaluation expertise as well as topical and regional expertise, reflecting the focal points of a programme and all stakeholders.

Wider benefits and opportunities of participatory, real-time evaluations

Despite the challenges discussed above, the evaluation of the AII provided important opportunities and benefits for capacity-building efforts with potential implications for future evaluations. This includes learning gains focussed on personal development (e.g. enhanced evaluation skills, better awareness of working cultures in different research capacity-building environments); formative learning to strengthen the implementation of capacity-building interventions, and contributions to the development of professional networks from a collaborative evaluation experience. These types of benefits resonate with wider literature on the *process use* of evaluations which emphasise the utility for stakeholders of being involved in the planning and implementation of evaluation (Patton, 1998; Forss et al., 2001).

First, the approach helped strengthen local evaluation capacity at some participating institutions. There is now an established group of people who are familiar with and know how to engage with real-time, theory- driven evaluations of highly complex interventions. We of course did not begin this evaluation in a vacuum of evaluation capacity on-the-ground, but it varied across consortia. Drawing on the experience of this particular evaluation, the external evaluators and consortia

Directors have concluded that, especially in lead institutions within consortia, evaluation officers and programme managers could conduct significant parts of a similar future evaluation relatively independently, given appropriate time and resources. This reflects wider learning on the value of self-assessment and participatory evaluations (e.g. Lusthaus et al., 1999; Horton et al., 2001). Trained evaluation experts in Africa are in high-demand and the retention of these individuals may present a new 'science policy challenge'. Exploring options for spreading this capacity more widely (for example through 'training trainer') merits further consideration. Given the increase in evaluation capacity, we can envisage future evaluations where external evaluator roles are smaller, for example focussed on facilitating framework development, reviewing internally produced deliverables with an independent lens, and providing some ongoing consultative support.

Second, the approach contributed to knowledge management across the consortia. For example, the evaluation was an important way for the large consortia to keep track of what was going on across dispersed partner networks. Evidence from the evaluation also had formative value as it was frequently discussed and reflected on at consortia annual meetings. The evaluation also contributed to the development of better publication and grant tracking systems as well as data management improvements in some consortia, which contributed to strengthening research management practices in some of their participating institutions.

Third, the approach helped provide timely evidence and shared learning to increase chances of programme success. For example, it helped highlight where adaptations in consortia were needed, facilitating modifications in consortia processes. Some examples across the initiative included: introducing more management support for the Directorate over time; focusing more on research management training and institutional level capacity-building; reflecting on incentives for retaining and attracting students; and reflecting on the scale of support offered through studentships and fellowships. Evaluation evidence has also been useful for fundraising, with some consortia using data to showcase progress.

The evaluation was also a forum for cross-consortia learning about common challenges and ways of addressing them, and about good practice. As communicated by a consortium Director, the opportunity for cross-consortia linkage and exchange around evaluation insights helped to identify

examples of good practices being implemented elsewhere (e.g. in supervision, research governance, course design), provided an opportunity to think beyond one specific intervention or set of activities and helped create a sense of community to sustain the resolve to address common challenges. Interaction between different consortia and with evaluators was also important for modifying the evaluation approach over time, while keeping the overall objectives of the AII at the forefront..

Final reflections

This article focusses on a particularly complex evaluation and reflects on the use of a participatory, real-time evaluation approach. We hope to have surfaced some of the intricate issues that future evaluations of similar interventions might face. While these insights may not *resolve* evaluation challenges, we hope they will contribute in helping make some evaluations more effective, useful and manageable for all involved.

Across the stakeholders involved in this evaluation, there have been both shared views, as well as a variety of experiences and differences of perception and opinion – between the consortia and evaluators, between different consortia, between evaluators, consortia and the funder. As discussed, some of the most challenging differences applied to issues such as the suitability of resources available for implementing the evaluation and the extent to which there was sufficient time for face-to-face engagement. Reconsidering the priority given to specific questions, collective success in securing additional finance for consortia evaluation officers for coordination activities, and efforts to embed additional consultation and interaction time over the course of the evaluation were particularly important for managing many challenges encountered. There are of course alternatives to participatory, real-time evaluations. While they are often more appropriate for less complex interventions and/or once initiatives are more mature, they should not be dismissed upfront. For example, if a funder has a specific interest in a clearly articulated and limited number of aspects of a more complex intervention (e.g. how a specific training programme is working within a wider framework of research capacity-building efforts, or what impact the complex intervention is having on a specific theme such as supervision or inter-institutional networking), a simpler and more *thematic evaluation* approach might be helpful as well Similarly, approaches rooted in benchmarking or impact tracing survey tools (Grant et al., 2010) might be possible if the

interest is in capturing outputs and impacts and if the emphasis is more on comparability than on learning about process variables and pathways to impact.

It is now widely accepted that different evaluation approaches all have their merits and limitations, and are – to varying degrees – fit for different purposes (see Stern et al., 2012). However, if evaluations aim to inform the policies and strategies of funders, researchers, policymakers and practitioners, it is essential for outcome evaluation and explanation to go hand-in-hand, so that decision-makers can understand both 'the why' and 'the how' of different choices and can use this learning in future decision-making.

Thematic evaluations centred on a limited set of issues are possibly easier to implement. However, they are likely to paint a less rounded picture, and may be more valuable and feasible at later stages in the development of complex interventions, than in the early stages when there are many unknowns and much uncertainty. Indeed, one of the key values of real-time, theory- - driven evaluations is in surfacing important issues. This is however also often a contributionmost difficult to obtain stakeholder and funder buy-in.

It is also essential not to lose sight of causality in thematic evaluations, and to remain focussed on the intervention's contribution story, avoiding overgeneralisation (Mayne, 2012).

An approach focussed more on *self-evaluation* is another alternative (see Horton et al., 2007). Here an external evaluator has more of an oversight, advisory and scrutiny role. The evaluator can help the programme practitioners establish their own theories of change, evaluation frameworks and templates, provide some training, and review self-produced evaluation and learning outputs to ensure sufficient attention to causality and appropriate backing by rigorous evidence. While these approaches are potentially less resource intensive, they require sufficient evaluation capacity from all stakeholders from the beginning and more resources devoted to evaluation within the intervention itself.

Lastly, evaluation is both a science and an art. Efforts to build effective evaluation capacity and to develop evaluation professionals should pay due consideration to the non-technical skills that are needed. Regardless of the analytical approach taken in an evaluation, there are a range of skills which evaluators need in order to effectively design and implement evaluations of complex interventions in international development and similar contexts.

These skills transcend topic-specific skill-sets, and include soft skills such as communication, negotiation, empathy, flexibility, cross-cultural awareness, responsiveness to criticism and ability to solicit, give and act on feedback. Such skills are particularly important to keep at the forefront of practice in participatory evaluations taking place in cross-cultural contexts, given the diversity of associated working cultures, values and behavioural norms. As such, they should be considered in the context of training the future cadre of evaluation practitioners and in future reflections on the reality of implementing different types of evaluations.

Conflict of interest

None

Acknowledgements

Notes

[Sage: add footnotes here]

References

Bamberger M, Vaessen J and Raimondo E (2015) *Dealing with complexity in development evaluation: a practical approach*. Thousand Oaks: Sage Publications.

Bates I, Akoto AYO, Ansong D, Karikari P, Bedu-Addo G, Critchley J, Agbenyega T and Nsiah-Asare A (2006) Evaluating health research capacity building: an evidence-based tool. *PLoS Med* 3(8): e299.

Bates I, Taegtmeyer M, Squire SB, Ansong D, Nhlema-Simwaka B, Baba A and Theobald S (2011) Indicators of sustainable capacity-building for health research: analysis of four African case studies. *Health Research Policy and Systems* 9: 14.

Bates I, Boyd A, Aslanyan G, and Cole DC (2014) Tackling the tensions in evaluating capacity strengthening for health research in low-and middle-income countries. *Health Policy and Planning* 30(3):334-344.

Bonfoh B, Raso G, Koné I, Dao D, Girardin O, Cissé G, Zinsstag J, Utzinger J and Tanner M (2011) Research in a war zone. Nature 474(7353): 569-571.

Boyd A, Cole DC, Cho DB, Aslanyan G and Bates I (2013) Frameworks for evaluating health research capacity strengthening: a qualitative study. *Health Research Policy and Systems* 11: 46.

Campbell N, Murray E, Emery J, Farmer A, Griffiths F, Guthrie B, Lester H, Wilson P and Kinmouth AL (2007) Designing and evaluating complex interventions to improve healthcare. *British Medical Journal* 334(7591): 455–9.

Chouinard JA (2014) Understanding relationships in culturally complex evaluation contexts. *Evaluation* 20(3): 332-347.

Chouinard JA and Cousins JB (2013) Participatory evaluation for development: Examining research-based knowledge from within the African context. *African Evaluation Journal* 1(1): 1-9.

Chu KM, Jayaraman S, Kyamanywa P and Ntakiyiruta G (2014) Building Research Capacity in Africa: Equity and Global Health Collaborations. *PLoS Med* 11(3): e1001612.

Cochrane G, Robin E, Marjanovic S, Diepeveen S, Hanlin R, Kryl D, Retter L, Yaqub O and Chataway J (2014) *The African Institutions Initiative: Insights from the First Four Years*. Santa Monica, CA: RAND Corporation.

Cole DC, Boyd A, Aslanyan G and Bates I (2014) Indicators for tracking programmes to strengthen health research capacity in lower-and middle-income countries: a qualitative synthesis. *Health Research Policy and Systems* 12:17.

Cooke J (2005) A framework to evaluate research capacity-building in health care. *BMC Family practice* 6(1): 44.

Connell JP and Kubish AC (1998) *Applying a theory of change approach to the evaluation of comprehensive community initiatives: progress, prospects, and problems*. The Aspen Institute. 2:15-44.

Cousins JB and Chouinard JA (2012) *Participatory evaluation up close: An integration of research-based knowledge*. Charlotte, NC: Information Age.

Daigneault PM and Jacob S (2009) Toward accurate measurement of participation: Rethinking the conceptualization and operationalization of participatory evaluation. *American Journal of Evaluation* 30(3): 330-348.

Ezeh A, Izugbara CO, Kabiru CW, Fonn S, Kahn K, Manderson L, Undieh AS, Omigbodun A, and Thorogood M (2010) Building capacity for public and population health research in Africa: the consortium for advanced research training in Africa (CARTA) model. *Global Health Action* 5:5693.

Forss K., Rebien CC and Carlsson J (2002) Process use of evaluations types of use that precede lessons learned and feedback. *Evaluation* 8(1): 29-45.

Gadsby EW (2011) Research capacity strengthening: donor approaches to improving and assessing its impact in low-and middle-income countries. *The international journal of health planning and management* 26(1): 89-106.

Glouberman S and Zimmerman B (2002) *Complicated and complex systems: what would successful reform of Medicare look like?* Toronto: Commission on the Future of Health Care in Canada.

Grant J, Brutscher PB, Guthrie S, Butler L and Wooding S (2010) *Capturing Research Impacts: A review of international practice.* Santa Monica, CA: RAND Corporation.

Horton D, Alexaki A, Bennett-Lartey S, Noele Brice K, Campilan D, Carden F, de Souza Silva J, Duong LT, Khadar I, Maestrey Boza A, Muniruzzaman IK, Perez J, Chang MS, Vernooy R and Watts J. (2003) Evaluating capacity development: experiences from research and development organizations around the world. Ottawa: IDRC.

Horton D and Mackay R (2003) Using evaluation to enhance institutional learning and change: recent experiences with agricultural research and development. *Agricultural Systems* 78(2): 127-142.

Horton D, Galleno V and Mackay R (2007) *Evaluation, Learning and Change in Research and Development Organizations: Concepts, Experiences, and Implications for the CGIAR.* ILAC Working Paper 2. Rome: Institutional Learning and Change (ILAC) Initiative.

Jones N, Bailey M and Lyytikainen M (2007). Research capacity strengthening in Africa: trends, gaps and opportunities. *A scoping study commissioned by DFID on behalf of IFORD*. London: ODI.

Ling T (2012) Evaluating complex and unfolding interventions in real time. *Evaluation* 18(1): 79-91

Lusthaus C, Adrien MH, Anderson G and Carden F (1999) *Enhancing Organisational performance. A toolbox for self assessment*. Ottawa: IDRC.

Marjanovic S, Hanney S and Wooding S (2009) *A historical reflection on research evaluation studies, their recurrent themes and challenges*. Santa Monica, CA: RAND Corporation.

Marjanovic S, Hanlin R, Diepeveen S and Chataway, J (2013) Research Capacity-Building In Africa: Networks, Institutions And Local Ownership. *Journal of International Development*, 25(7): 936-946.

Maselli D, Lys JA and Schmid J (2006) *Improving impacts of research partnerships. Swiss Commission for Research Partnerships with Developing Countries, KFPE.* Bern; Geographica Bernensia.

Mayne, J. (2012). Contribution analysis: Coming of age?. *Evaluation*, 18(3), 270-280.

Minja H, Nsanzabana C, Maure C, Hoffmann A, Rumisha S, Ogundahunsi O, Zicker F, Tanner M and Launois P (2011) Impact of Health Research Capacity Strengthening in Low- and Middle-Income Countries: The Case of WHO/TDR Programmes. *PLoS Neglected Tropical Diseases.* 5(10):e1351.

Neilson S and Lusthaus C (2007) *IDRC-Supported capacity building: developing a framework for capturing capacity changes.* Ottawa: International Development Research Centre. Available at: https://idl-bnc.idrc.ca/dspace/bitstream/10625/29146/1/125252.pdf (accessed 10 June 2016).

NORAD (2009) *Evaluation of the Norwegian Programme for Development, Research and Education (NUFU) and of Norad's Programme for Master Studies (NOMA).* Oslo: NORAD. Available at: http://evalueringsportalen.no/evaluering/evaluation-of-the-norwegian-programme-for-development%252C-research-and-education-%2528nufu%2529-and-of-norad%25E2%2580%2599s-programme-for-master-studies-%2528noma%2529/EvaluationReport_7_09.pdf/@@inline (accessed 10 June 2016).

Patton MQ (1998) Discovering process use. *Evaluation* 4(2): 225-233.

Patton MQ (2001) Evaluation, knowledge management, best practices, and high quality lessons learned. *The American journal of evaluation* 22(3): 329-336.

Pawson R and Tilley N (1997) *Realist Evaluation.* London: Sage.

Rogers P (2008) Using programme theory to evaluate complicated and complex aspects of interventions. *Evaluation* 14(1): 29–48.

SIDA (2005) *Evaluation of HEPNet in SSA.* Stockholm: Sida. Available at: http://www.sida.se/contentassets/12bf5529dcd545a195bec06e7ff1a07d/evaluation-of-hepnet-in-ssa_628.pdf (accessed 10 June 2016).

Stern E, Stame N, Mayne J, Forss K, Davies R and Befani B (2012) *Broadening the range of designs and methods for impact evaluations. Report of a study commissioned by the Department for International Development*. London: Department for International Development.

Taut S (2007) Studying self-evaluation capacity-building in a large international development organization. *American Journal of Evaluation* 28(1): 45-59.

Trostle J (1992) Research capacity-building in international health: definitions, evaluations and strategies for success. *Social science & medicine* 35(11): 1321-1324.

Vasquez EE, Hirsch JS, Giang LM and Parker RG (2013) Rethinking health research capacity strengthening. *Global public health* 8(sup1): S104–S124.

Weiss C (1995) Nothing as Practical as Good Theory: Exploring Theory-based Evaluation for Comprehensive Community Initiatives for Children and Families. In Connell et al. (eds) *New Approaches to Evaluating Community Initiatives: Concepts, Methods, and Contexts*. Washington DC: Aspen Institute. 65-92.

Whitworth JA, Kokwaro G, Kinyanjui S, Snewin VA, Tanner M, Walport M and Sewankambo N (2008) Strengthening capacity for health research in Africa. *The Lancet* 372(9649): 1590-1593.

---

[1] For more information see: http://www.wellcome.ac.uk/aii
[2] For discussions on this point related to the initiative, see Marjanovic et al (2013) and Cochrane et al (2014)
[3] Enhancing Support for Strengthening the Effectiveness of National Capacity Efforts – http://www.who.int/tdr/partnerships/initiatives/essence/en/
[4] The baseline assessments considered capacity in terms of researchers at different stages of the career pathway; existing postgraduate training programmes at institutions; continual professional development opportunities; research governance, management and administration capacity at partner institutions; the research funding environment; physical and ICT infrastructure capacities; and baseline collaborative networks across different stakeholder groups (e.g. academic, policy, funder, practitioner).
[5] Through a journal article: Marjanovic et al. (2013)
[6] The American Society of Tropical Medicine and Hygiene (ASTMH), 59th Annual Meeting, Atlanta, Georgia, 4 November 2010 and the London International Development Centre (LIDC), Measuring Impact of Higher Education for Development Conference, London, 19 March 2013. Slides available at: http://www.lidc.org.uk/_assets/Marjanovic.pdf
[7] THRiVE Evaluation and Learning Workshop, Makerere University, Kampala, Uganda - 13th July 2013
[8] Cochrane et al. (2014)