

Applying evaluative thinking to value for money: The Pakistan Sub-national Governance Programme

Julian King and Stephanie Allan

Value for Money (VfM) poses an evaluative question about how well resources are used, and whether the resource use is justified (King, 2017). In international development there is increasing scrutiny on VfM. Currently, however, VfM assessment in the sector leaves substantial room for improvement. For example, VfM frameworks often comprise a collection of indicators of variable quality, devoid of an explicitly evaluative judgement. This article demonstrates the use of explicit evaluative reasoning in VfM assessment, using rubrics tailored to the programme context to support mixed methods evaluation of VfM. This approach facilitates transparent and well-reasoned judgements that respond to donor requirements for information that supports accountability and good resource allocation, as well as learning and adaptive management. The example of the Sub-national Governance (SNG) Programme in Pakistan (supporting reforms in public financial management, planning, and innovative service improvement pilots) is used to illustrate the use of this VfM approach in addressing the challenges that arise when evaluating VfM in complex, adaptive programmes. Core principles of the approach are transferable to other sectors.

Introduction

In international development, because of limited aid budgets and political pressures to be accountable for the use of taxpayers' funds, it is accepted that aid should be well targeted and managed effectively (Adou, 2016). These drivers have led to an increased interest in VfM over the past 15 years (Fleming, 2013).

There are multiple definitions of VfM (King, 2017). In programme evaluation, the term has tended to be used synonymously with the economic concept of efficiency (e.g., Schwandt, 2015), with economic methods of evaluation, and cost-benefit analysis in particular, widely considered the gold standard for its measurement (Julnes, 2012). Governmental, philanthropic, and international development publications have defined VfM in various ways that touch on maximising value in the procurement of inputs; delivery of outputs; achievement of outcomes; relationships between resources and outcomes (e.g., return on investment); and/or impacts on equity (King, 2017). For example, the UK's Department for International Development (DFID) defines VfM as "maximising the impact of each pound spent to improve poor people's lives" and further breaks this concept into five dimensions, labelled economy, efficiency, effectiveness, cost-effectiveness, and equity (DFID, 2011).

This article builds on six propositions from King (2017): i) that VfM is an evaluative question about an economic problem; ii) that VfM questions should be addressed through explicit evaluative reasoning; iii) that economic methods of evaluation (such as cost-benefit analysis, cost-effectiveness analysis, and cost-utility analysis) are approaches to evaluative reasoning (though perhaps not labelled as such in economic texts); iv) that economic methods of evaluation are relevant and powerful for VfM assessment; v) that economic methods of evaluation alone may be insufficient to fully answer an evaluative

question about VfM; and vi) that in cases where economic methods of evaluation can enhance VfM assessment but are insufficient on their own, economic evaluation could be incorporated within a mixed methods assessment of VfM—for example, results from a cost-benefit analysis could be combined with evidence gathered using other methods (quantitative and/or qualitative), tailored to context, guided by an overarching process of evaluative reasoning. Table 1 provides further detail on these propositions.

Table 1. VfM propositions from King (2017)

Proposition	Rationale
VfM is “an evaluative question [concerned with merit, worth or significance] about an economic problem [resource allocation]” (King, 2017, p. 102).	VfM may be a broader concept than efficiency or worth—e.g., the inclusion of equity in DFID’s framework recognises that “reaching marginalised groups may entail additional effort and cost” (ICAI, 2018, p. i). There may be further dimensions to good resource allocation beyond efficiency and equity, such as deontological ethics (Adler & Posner, 2006), relevance and sustainability (King & OPM, 2018). Balancing multiple criteria can involve trade-offs, suggesting that good resource allocation “is a matter of context and perspective” (King, 2017, p. 102).
Evaluative questions about VfM should be addressed through explicit evaluative reasoning (Yarborough, Shulha, Hopson, & Caruthers, 2011).	Evaluation has been defined as “the process of determining the merit, worth or significance of something” (Scriven, 1991, p. 139). This process “does not aim simply to describe some state of affairs but to offer a considered and reasoned judgement about that state of affairs” (Schwandt, 2015, p. 47). “Making a reasoned choice about value and being able to defend it is what distinguishes evaluative thinking from [other] critical thinking” (Vo, Schreiber, & Martin (2018, p. 40). The “fundamental problem” in this endeavour is “how one can get from scientifically supported premises to evaluative conclusions” (Scriven, 1995, p. 51). “Explicit evaluative reasoning provides the means to make robust judgements from evidence” (King & OPM, 2018, p. 9).
Economic methods of evaluation are approaches to evaluative reasoning.	Economic methods of evaluation (e.g., cost-benefit analysis, cost-effectiveness analysis, cost-utility analysis) involve systematically identifying, measuring, valuing, and comparing the costs and consequences of alternative courses of action (Drummond, Sculpher, Torrance, O’Brien, & Stoddard, 2005) and synthesising the evidence to make a judgement. The methods can be viewed as specific formulations of an approach to evaluative reasoning described as “numerical weight and sum” (Scriven, 1991). They principally yield estimates of efficiency (e.g., net present value, incremental cost-effectiveness ratio) though wider considerations such as equity may also be discussed (Drummond et al., 2005).

Proposition	Rationale
Economic methods of evaluation are relevant and powerful for VfM assessment.	<i>For example</i> , by: providing a framework for systematic analysis of both costs and consequences; focusing explicitly on the opportunity cost of resource use; valuing costs and consequences in commensurable units (in the case of cost-benefit analysis); explicitly considering the relative timing of costs and consequences; and facilitating insights about risk and uncertainty through scenario and sensitivity analysis (King, 2017).
Economic methods of evaluation alone may be insufficient to fully answer an evaluative question about VfM.	<i>For example</i> , cost-benefit analysis: is primarily concerned with efficiency (Adler & Posner, 2006) whereas VfM may be multi-criterial (King, 2017); takes a consequentialist perspective, whereas programme processes may have value independent from the outcomes they produce (Julnes, 2012); may obscure qualitative differences in the perspectives of different groups (e.g., power imbalances, diverging worldviews or interests) (Julnes, 2012); may in practice leave out intangible dimensions of value because they are too hard to estimate given constraints of time, evaluation resources, analyst capability or current techniques (Adler & Posner, 2006). Moreover, though normative values and assumptions from welfare economics may be widely accepted in economics and policy analysis, there could be circumstances where we do not take for granted, for example, that individuals are the best judges of their own welfare, that individuals have preferences that are stable over time, that utility increases when individual preferences are satisfied, that total utility should comprise the aggregate utilities of each individual, or that any net gain in overall efficiency (the size of the pie) is worthwhile regardless of its impact on equity (how the pie is shared) as long as the winners could <i>potentially</i> compensate the losers (Adler & Posner, 2006; Drummond et al., 2005).
Where economic methods of evaluation can enhance VfM assessment but are insufficient on their own, an economic evaluation could be incorporated within a mixed methods assessment of VfM.	<i>For example</i> , this approach could retain many of the strengths of economic evaluation while providing additional flexibility to accommodate: holistic or analytic evaluation, using absolute (grading, rating, scoring) or relative (ranking, apportioning) systems; and mixed methods assessment drawing together diverse streams of evidence. Such an approach could be used, for example, to transparently: balance multiple criteria; assess “process value” independently from outcomes; accommodate various normative positions on equity; contrast and support deliberation on qualitative differences between groups of people; and support robust qualitative assessment of intangible values that are hard to measure and monetise (King, 2017).

Integrating an economic evaluation within a wider, mixed methods VfM assessment would require an overarching system of evaluative reasoning to synthesise and make judgements from diverse streams of evidence. Although there are multiple ways to approach evaluative reasoning (e.g., see Schwandt, 2015) a common model comprises four steps of: i) establishing criteria of merit, worth, or

significance—dimensions of performance that are relevant and important to an evaluative judgement; ii) defining standards for each criterion, specifying “what the evidence would look like at different levels of performance” (Davidson, 2014); iii) gathering and analysing evidence of performance against the standards; and iv) synthesising the results into an overall judgement (Fournier, 1995). Evaluation rubrics (Davidson, 2005) offer a practical approach to support this process (King, McKegg, Oakden, & Wehipeihana, 2013).

King and OPM (2018) described an approach to operationalising King’s (2017) theoretical model. This article presents an example of what the approach looks like in practice, underpinned by the use of rubrics to support transparent judgements about VfM from mixed methods evidence. It starts by presenting a challenge: the need to assess VfM in a complex, adaptive governance programme and the specific contextual issues that presented challenges for evaluating VfM. It then describes the approach taken, the framework that was developed, the evidence used, and in broad terms the conclusions reached. It concludes by discussing the strengths, limitations, and potential of the approach.

The Sub-national Governance (SNG) Programme

The Sub-national Governance (SNG) Programme supported provincial and local governments in two provinces in Pakistan (Khyber Pakhtunkhwa (KP) and Punjab) to improve the delivery of basic services. Financed by DFID, and managed by Oxford Policy Management Ltd. (OPM), the programme operated from 2013 to March 2018 in 12 districts, six each in KP and Punjab, supporting reforms in public financial management, governance, and planning, and operating a challenge fund to finance innovative service improvement pilot projects with the view that successful pilots would be adopted and scaled up by the provincial governments.

Collectively, these diverse work streams aimed to deliver results across three main areas: decisions by government based on robust evidence; public services that are more responsive to people's needs; and strengthened government capability to deliver basic services. The intended outcome of the programme was that “poor people in Punjab and Khyber Pakhtunkhwa report that services are better meeting their needs”, and its intended impact was “a more stable democracy in Pakistan, through increased trust in Government” (Department for International Development (DFID) Pakistan, 2012, p. 20).

A challenging context in which to evaluate VfM

SNG was a complex governance reform programme. Its complexity derived in part from its operating environment, working with sub-national governments in Pakistan. This required a keen understanding of the prevailing political economy in Punjab and KP, including the volatile security situation, the dynamic power structures within the two provincial governments, and their changeable relationships with central and district governments during a period of considerable decentralisation reform. The VfM assessment needed to recognise this context and assess the programme's achievements relative to what could reasonably be expected in the complex and rapidly changing environment. This context meant that some programme outcomes were difficult to measure and that some quantitative indicators may be ambiguous to interpret. There were no meaningful external benchmarks against which SNG's performance could be compared, nor could VfM assessment rely on “off the shelf” definitions of VfM, such as efficiency being the cost per unit of outputs or outcomes. This meant that a VfM assessment had to define performance standards from scratch.

Furthermore, there was no linear relationship between SNG's outputs and its outcomes; other factors and forces were present which

meant that the delivery of outputs would not necessarily translate into a proportional change at outcome level; and, when outcome-level changes were registered, attribution of these (in full or in part) to the programme was problematic.

An additional challenge in assessing VfM of SNG was that the value of the programme lay not only in the achievement of the specified outcomes, but also in its capacity to generate learning to influence other relevant programmes, and its ability to be adaptive, responding to lessons learnt as well as emergent opportunities and challenges. A sophisticated VfM framework was needed to systematically capture these learnings and the emergence of adaptive programming. It also needed to not unduly penalise the programme if some of its interventions were found to be ineffective and were subsequently disbanded, as this was inherent to the adaptive design. Finally, the flexible and responsive nature of SNG meant that many of the programme's interventions and results could not be defined at the outset, hence the benchmarks against which to assess performance also needed to be sufficiently flexible.

Institutional context

As is routine for all DFID programmes, the SNG Programme underwent annual external reviews led by DFID, which included an assessment of VfM. Under DFID's annual review process, it was expected that the contractor (OPM) would furnish all necessary evidence of performance and VfM, together with suggested ratings. DFID's annual review teams scrutinised the evidence and made the final judgements about the programme's performance and VfM. For the first three reviews, the assessment focused primarily on the management of inputs, as the longer term results of the programme had yet to be registered. In 2016, as the programme entered its fourth year of operation, DFID called for the VfM framework to be reviewed

and expanded to include analysis of outcomes.

VfM in international development has typically not been assessed using explicit evaluative reasoning, except in cases where economic methods of evaluation have been applied. Donor requirements continue to include the use of linear project management, monitoring, and accountability frameworks such as the Logical Framework (or logframe) (Schiere, 2016). VfM assessments have similarly tended to follow a reductionist, indicator-based approach. There is a risk of making invalid assessments of VfM if tied to a narrow set of indicators—for example, by focusing on activities that are easy to measure but relatively unimportant, or by focusing on quantification of outputs and outcomes at the expense of more nuanced consideration of their quality and value (King & Guimaraes, 2016). Moreover, such methods struggle to accommodate the adaptive and incremental approaches which are commonly used on reform programmes to gain traction (Andrews, 2013; Archibald, Sharrock, Buckley, & Young, 2018).

Cost-benefit analysis of DFID's aid investments is often applied prospectively at the programme appraisal stage to assess its potential worth but may be problematic to use for VfM appraisal during and immediately after the delivery of a programme—in the SNG instance, because of problems assigning monetary valuations to intermediary outcomes (such as institutional strengthening of provincial government departments), as well as the need to consider not only efficiency but also equity.

Approach and methods

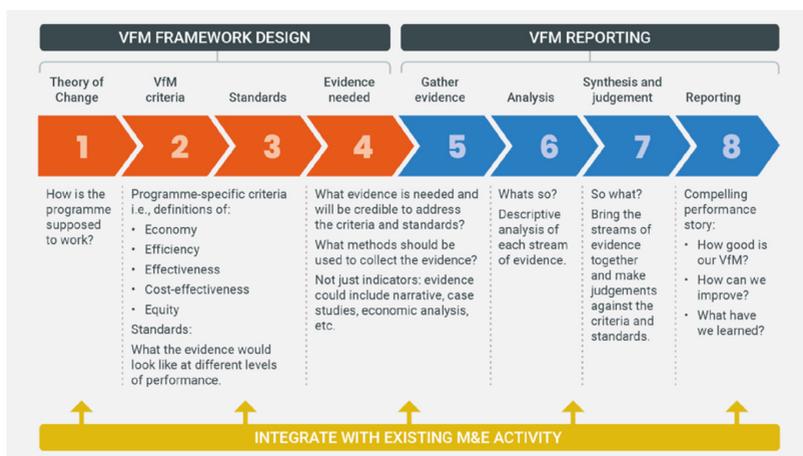
The key innovation introduced in the VfM assessment of the SNG Programme was the use of rubrics to guide evaluative reasoning. Using the approach described by King and OPM (2018), the evaluators developed rubrics for DFID's VfM criteria (economy, efficiency,

effectiveness, cost-effectiveness, and equity), tailored to the specific activities, outputs, and outcomes of the programme. Mixed methods evidence (qualitative, quantitative, and economic) was then gathered, analysed, and synthesised as detailed in the following sections.

Staged approach to explicit evaluative reasoning

The VfM evaluation was carried out in two distinct stages: VfM framework design, and VfM reporting. Each stage involved a 1-week meeting between the SNG teams and the authors. The VfM framework was developed during August–September 2016 and approved by DFID in October 2016. The first VfM assessment was carried out in February–March 2017 and the second in February–March 2018. Figure 1 summarises the steps in the process (King & OPM, 2018), which are explained sequentially in the following paragraphs.

Figure 1. Stages in VfM Evaluation



Source: King and OPM (2018, p. 21)

Theory of change: A theory of change “explains how activities are understood to produce a series of results that contribute to achieving

the final intended impacts” (Rogers, 2014, p. 1). By making prior evidence and assumptions about the intervention explicit, a theory of change helps to facilitate clear thinking about the intended inputs, activities, outputs, and outcomes of the intervention. Because it describes the intended operation of the programme in systematic terms, a theory of change provides a reference point for considering criteria of merit, worth, or significance (Davidson, 2005).

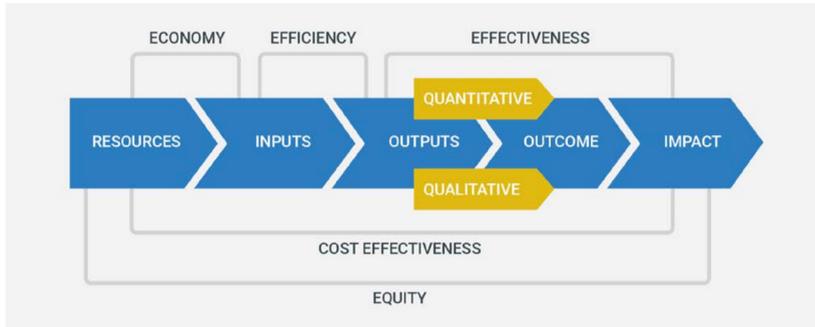
In SNG, the intended outcome had been defined at a high level (“services better meet the needs of poor people in Punjab and KP”)—leaving a relatively long conceptual journey from outputs to the intended outcome. The first step in developing a VfM framework for the programme was to review the theory of change, in consultation with SNG monitoring and evaluation (M&E) advisers, technical advisers, and management, to articulate the implied causal pathways between outputs and final outcomes. Through this process, a chain of intermediary outcomes was elucidated.

In particular, a key population-level intermediary outcome was identified, connecting changes within governments to the final outcome and impact indicators: “*increased funding and/or efficiency improvements for service delivery, targeted to needs*”. It was reasoned that the changes within governments, stemming from SNG, should result in additional resources for service delivery, targeted to needs, ultimately resulting in increased resources utilised by districts and service delivery units, and thus impacting on beneficiaries.

Criteria and standards: Criteria are selected dimensions of performance that describe at a broad level, the aspects of performance that need to be evidenced to support an evaluative judgement about VfM. When evaluating VfM in DFID-funded programmes, the core criteria to be used are specified in the document, *DFID’s Approach to Value for Money* (DFID, 2011) and are illustrated in Figure 2. These should not be regarded as universal VfM criteria; however, the use of

DFID's criteria was a requirement for the evaluation discussed here.

Figure 2. DFID's VfM Criteria



Source: King and OPM (2018, p. 11)

Economy, efficiency, and effectiveness, as defined by DFID (2011) divide the concept of VfM into discrete sections of a programme's theory of change. Economy focuses on the cost and value of inputs. Efficiency focuses on the transformation of inputs, by sets of activities, into outputs. Effectiveness focuses on the achievement of outcomes. Cost-effectiveness looks at the whole results chain from inputs through to impact. Although this approach offers a pragmatic breakdown of VfM criteria, the simplified depiction of a results chain may be insufficiently nuanced to reflect contemporary challenges in governance interventions such as adaptive programming.

DFID's (2011) definitions can also be challenging to apply in practice because they are expressed at a broad conceptual level that does not necessarily link in an intuitive way to the inputs, outputs, and outcomes of a specific programme or context. Furthermore, these definitions do not provide a transparent basis for distinguishing "good" VfM from "excellent" or "poor" VfM. ICAI (2011) offers a suggested set of standards through a traffic light rating system, but these, like DFID's criteria, are generic (King & OPM, 2018).

For SNG, programme-specific definitions of the criteria were developed, together with more detailed sub-criteria. Table 2 summarises DFID’s generic question for each criterion, alongside the SNG-specific criteria and sub-criteria that were developed. Performance standards were also developed (Table 3), defining levels of performance for each criterion. Although these levels can be labelled in various ways (e.g., from “excellent” to “poor”), in this instance a visual “traffic light” system was adopted.

The criteria and standards were developed in a series of workshops with SNG M&E advisers, technical advisers, and management, by identifying and elaborating on key inputs, processes, outputs, and outcomes from the theory of change. This collaborative process was important because it facilitated a nuanced understanding of context, political economy, adaptations, and real-world functioning. The process of co-developing criteria and standards surfaced a range of values and perspectives among the different actors and provided a forum to incorporate and reconcile these within the more formally documented expectations (King et al., 2013). As this VfM evaluation was being undertaken for a DFID programme, the criteria and standards were subsequently reviewed and signed off by DFID, providing assurance of their face validity for the core purpose of the VfM assessment.

Table 2. VfM Criteria used in the SNG Programme

VfM criteria	DFID's VfM questions	SNG-specific definition	Sub-criteria (overview)
Economy	"Are we or our agents buying inputs of the appropriate quality at the right price? (Inputs are things such as staff, consultants, raw materials and capital that are used to produce outputs)" (DFID, 2011, p. 4).	The SNG team manages programme resources economically, buying inputs of the appropriate quality at the right price.	Average unit costs of consultants (national/ international) compared to contract; trends in average costs of significant items compared to contract; supplier negotiation and contract management ensuring inputs remain cost-competitive; one-off cost savings secured through negotiation.
Efficiency	"How well do we or our agents convert inputs into outputs? (Outputs are results delivered by us or our agents to an external party. We or our agents exercise strong control over the quality and quantity of outputs)" (DFID, 2011, p. 4).	The SNG Programme produces the intended quality and quantity of deliverables, within the available resources.	Comparison of actual delivery against work plans. "Deliverables" refers to completion of planned activities and delivery of "products" such as draft legislation, rules and guidelines, and training. Recognition of programme activities that varied from the work plan to pursue new opportunities as they arose and drop activities without traction. In this fixed budget environment, and bearing in mind the bespoke, heterogeneous nature of deliverables and consequent lack of benchmarks, cost-output ratios were not used.

VfM criteria	DFID's VfM questions	SNG-specific definition	Sub-criteria (overview)
Effectiveness	"How well are the outputs from an intervention achieving the desired outcome on poverty reduction? (Note that in contrast with outputs, we or our agents do not exercise direct control over outcomes)" (DFID, 2011, p. 4).	The SNG Programme achieves its intended changes in public financial management, governance, and planning systems, and service improvement pilots, in Punjab and Khyber Pakhtunkhwa.	"Intended changes" were further specified in sub-criteria corresponding to each outcome area specified in the logframe (e.g., "improved Open Budget Index (OBI) score", "district budgets based on evidence of people's needs", "strengthened capacity of local governments to deliver services", "government uses geographic information systems (GIS) to monitor and plan quantity and quality of services", inter alia). These were individualised to each province, given the different issues and strategies at play.
Cost-effectiveness	"How much impact on poverty reduction does an intervention achieve relative to the inputs that we or our agents invest in it?" (DFID, 2011, p. 4).	The SNG Programme contributes to increased funding and/or efficiency improvements for services to meet identified needs in Punjab and Khyber Pakhtunkhwa.	Ratio of fiscal value created (increased funding allocations and utilisation to education and health services, plus efficiency gains) to fiscal value consumed (resources invested in the SNG Programme).
Equity	"When we make judgements on the effectiveness of an intervention, we need to consider issues of equity. This includes making sure our development results are targeted at the poorest and include sufficient targeting of women and girls" (DFID, 2011, p. 3).	Changes in needs-based planning and resource allocation contribute to reducing inequities by targeting resources to poor people, women, and girls.	Proportion of increased funding allocated to, and utilised by services for poor people, women, and girls.

Table 3. VfM Standards used in the SNG Programme

					
Economy	Substantially exceeded expectation	Moderately exceeded expectation	Met expectation	Moderately did not meet expectation	Substantially did not meet expectation
Efficiency	SNG deliverables for the year substantially exceeded work plan and in line with allocated budget.	SNG deliverables for the year moderately exceeded work plan and in line with allocated budget.	SNG deliverables for the year completed according to work plan and in line with allocated budget.	SNG deliverables for the year moderately did not meet work plan and/or moderately exceeded budget.	SNG deliverables for the year substantially did not meet work plan and/or substantially exceeded budget.
Effectiveness	Substantially exceeded expectation.	Moderately exceeded expectation.	Met expectation.	Moderately did not meet expectation.	Substantially did not meet expectation.
Cost-effectiveness	Increased funding utilisation and efficiency gains exceed combined DFID resourcing for SNG and consequential provincial and district government investments in governance, planning, and public financial management reform.	Increased funding utilisation and efficiency gains exceed DFID resourcing for SNG Programme.	Increased funding allocation and efficiency gains exceed DFID resourcing for SNG Programme.	Funding allocation for services moderately below DFID resourcing for SNG Programme.	Funding allocation for services substantially below DFID resourcing for SNG Programme.

					
Equity	A substantial proportion of increased funding is utilised by services for poor people, women, and girls.	A moderate proportion of increased funding is utilised by services for poor people, women, and girls.	Needs-based planning and resource allocation includes explicit targeting of services for poor people, women, and girls.	Needs-based planning and resource allocation includes implicit targeting of services for poor people, women, and girls.	Needs-based planning and resource allocation does not target services for poor people, women, and girls.

Determining evidence requirements: In a logical, sequential process of evaluation design, the theory of change, criteria, and standards provide a necessary foundation for determining what evidence is needed and will be credible to support evaluative judgements. After clarifying the criteria and standards, evidence that is relevant to those criteria and standards can be identified. This sequence of evaluation design helps ensure the validity of indicators and that the choice of evaluation methods is aligned with the context and values embedded in the programme (King et al., 2013).

Having developed performance standards, it became obvious that some of the evidence needed to support evaluative judgements was qualitative. Indicator-based measurement can make a valuable contribution to evaluating programme performance and VfM. Indicators alone, however, were insufficient to support evaluative judgements about this complex programme. Indicators by their very nature are narrow and provide individual pieces of measurable evidence that correlate with the VfM criteria. Broader contextual evidence was also needed, to provide further information about performance and support appropriate interpretation of the indicators (King & OPM, 2018).

In addition to general quantitative and qualitative evidence, economic evidence can play an important role in determining VfM (King, 2017)—and in SNG, a limited form of cost-benefit analysis was undertaken, with both costs and benefits being valued monetarily and analysed using a discounted cashflow model (DFID’s use of the term “cost-effectiveness” refers to a general principle of comparing inputs to impact and does not prescribe the use of “cost-effectiveness analysis” as the method of economic evaluation). Rationale for the cost-effectiveness criteria and standards presented in Tables 2 and 3 are summarised in Figure 3.

Figure 3. Assessing “Cost-effectiveness” in the SNG Programme

For SNG, “increased funding and/or efficiency improvements for services to meet identified needs” were used as proxy benefit measures, corresponding to the population-level intermediary outcome indicator in the theory of change. Costs were represented by the financial resources invested in the reforms by DFID and the provincial governments of Punjab and KP. The cost-effectiveness of the programme should ultimately be realised through improvements in democratic governance resulting from better access and satisfaction that should be achieved in the two provinces over the longer term. The defined cost-effectiveness criterion reflected the programme theory that increases in funding and efficiency improvements would be the practical result of improved allocative efficiency in public financial management and governance. These intended results would flow logically from governmental changes (e.g., increased provincial resources to finance services, needs-based planning and budgeting, improved flow of funds to districts), resulting initially in increased *allocations* to districts, and subsequently in increased *utilisation* of funds by health, education, and other services (which should in turn translate into more beneficiaries benefiting from more services, if those services are effective). The rationale for the use of these standards was that: i) in order to be worth the resources invested, the programme should create more value than it consumes; ii) a measurable and relevant near-term proxy for “value created” is the increased flow of funds to districts, according to the

rationale in the previous paragraph; iii) increased funding *allocation* to districts exceeding DFID's resourcing represents a reasonable minimum benchmark in that SNG's influence over allocations, though not direct, is somewhat closer than increases in funding *utilisation*; iv) increased *utilisation* of funding sets a higher bar and is ultimately more important than increases in allocations because it is closer to a real impact on access to services; and v) the total fiscal resources invested in governance, planning, and public financial management reforms in Punjab and KP included both DFID funds and resources invested by the provincial governments in implementing the reforms.

Gathering evidence: The evidence needed to address the criteria was collated and analysed by the SNG M&E advisers. Evidence was drawn from a range of sources including programme accounting data, provincial government budget tracking data, quarterly reports produced by the SNG teams, logframe ratings produced each year as part of DFID's annual review process, and a series of "output briefs" and case studies prepared by the SNG teams to templates provided by the evaluation team. The output briefs systematically addressed each output in the theory of change. These narrative briefs were produced to support accurate and well-evidenced evaluative judgements, with appropriate attention to context. Each output brief described: the problem to be addressed; the intervention, and institutional context; the most significant changes in government; the evidence that supported these claims; and assessment of the likely sustainability of these changes. They also documented the ways in which changes in government systems contributed to increased service delivery which was responsive to needs.

Making judgements about performance and VfM: Analysis, synthesis, and judgements were facilitated by the evaluation team in a

1-week workshop with the SNG M&E advisers, technical advisers, and management. *Analysis* involved examining each stream of evidence separately, and *synthesis* involved combining the components to reach a holistic understanding—for example, triangulating and considering the totality of evidence including any areas of corroboration or contradiction between evidence sources, and using qualitative evidence to support accurate interpretation of quantitative data (King et al., 2013). Judgements were then made against each set of criteria and standards individually.

Once the evidence and judgements for each criterion were in place, an overall judgement of VfM was made. This was done at the end of the assessment process, by bringing the teams back together for a process of weighting and judgement. Making the overall assessment required the evaluators and SNG teams to stand back and consider the relative importance of different criteria when making the overall judgement. For example, in the final year of the programme, cost-effectiveness and equity were considered the most important criteria as they related most closely to impacts for the populations of the two provinces.

Involving the SNG teams in conducting analysis, synthesis, and making judgements from the evidence positioned evaluative reasoning in the VfM assessment as “a collaborative, social practice” (Schwandt, 2018, p. 125). This was valuable for several reasons. First, given the complexity and evolving nature of the programme, it provided the opportunity to contextualise and validate judgements by accessing the team’s in-depth knowledge. Second, it provided for an inclusive, robust, and rigorous process of making judgements. Third, it provided a forum to identify and document key programme adaptations and lessons learnt that could inform future interventions of a similar nature. Fourth, it provided an opportunity for evaluation capacity building.

Contribution tracing: Determining the contribution of SNG to the outcomes seen was a particular challenge. The results of the programme are best conceptualised as the synergistic effect of interventions (e.g., ideas, influencing, training, resources, documents) produced by the SNG teams, and the adoption and spread of these and similar interventions by the provincial governments themselves. Additionally, other development programmes may contribute to the intended outcomes. In this context, and without the possibility of a measurable counterfactual, a theory-based approach called *contribution tracing* (Befani & Mayne, 2014) was used to assess the strength of causal claims.

Contribution tracing combines the approaches of contribution analysis (Mayne, 2008) and process tracing (Van Evera, 1997). It involves iterative analysis of causal claims along a detailed theory of change, with the insight that Bayesian inference can be applied to test evidence for causal links (Bayes, 1763; Befani & Stedman-Bryce, 2017). Overall, contribution tracing found it was possible to make a strong case that SNG activities had produced significant improvements in planning and budgeting processes but difficult to show that these improved planning and budgeting processes have yet improved service delivery.

Additionally, scenario analysis was undertaken to consider, at cost-effectiveness level, the proportion of increased allocations and efficiency gains that may be attributed to SNG. A framework of four economic considerations (deadweight, displacement, attribution, and drop-off), adapted from Social Return on Investment (Nicholls, Lawlor, Neitzert, & Goodspeed, 2012) was used to provide a systematic way of considering the relative contributions of the programme and external factors to outcomes, using transparent rationale. This framework facilitated critical appraisal and explanation of the ways in which the programme may contribute to the increase in health and education spending. This process yielded percentage estimates which were applied to the fiscal indicators, with sensitivity analysis.

Results

The purpose of this article is to describe the use of an approach to evaluative reasoning in VfM assessment, and not to present the full results of the SNG VfM assessment. Nevertheless, it is worth briefly outlining key findings. The 2017 VfM evaluation found that the SNG Programme achieved high ratings against the standards for economy, efficiency, effectiveness, cost-effectiveness, and equity—and therefore was providing strong VfM overall. For example, in the assessment of cost-effectiveness, analysis of budgetary trends led to the finding that the provincial governments' increased allocations to health and education, which could be directly linked to SNG interventions, already exceeded the investment in the programme by a substantial margin—with further allocations expected in future years. In the assessment of equity, it was found that the programme had addressed key service delivery issues affecting marginalised groups and that the increased spending on health and education was effectively targeted.

Discussion

It is good evaluation practice to consider VfM in any public investment—not only because of the opportunity cost (foregone alternatives) associated with allocating limited resources, but also because of political drivers to be transparent and accountable to taxpayers. In addition, VfM assessment can be treated as an opportunity to reflect, learn, adapt, and improve interventions. Though a governance example has been presented here, principles of the approach (King, 2017; King & OPM, 2018) are transferrable to other sectors.

The use of explicit evaluative reasoning to assess VfM facilitated transparent and valid judgements by linking diverse streams of evidence to the programme design and outcome logic. In so doing, it responded to donor requirements, as indicated in the following feedback:

As DFID works to build prosperity and stability in Pakistan, ensuring our investments represent good value for money is at the forefront of all programming decisions. The framework presented here offers a disciplined and pragmatic approach to analysing VfM in a hard-to-quantify sector. It represents a marked step forward in our monitoring, evaluation and learning framework, and is an important tool not only for DFID's own internal accountability, but also our accountability to those living in poverty in Pakistan. (Feedback to the evaluators from DFID Head Office)

Incorporating economic evaluation within a mixed methods assessment involves trade-offs in clarity. On one hand, it arguably weakens some benefits of economic evaluation as described in Table 1. For example, it promotes deliberative balancing of criteria rather than clean, numerical valuation and synthesis of costs and consequences into a net present value or other unifying indicator.

On the other hand, rubric-based VfM assessment promoted clarity in other ways: it accommodated a nuanced appraisal of the programme's responsiveness to the evolving context, alongside other important considerations including the targeting and value of changes in provincial resource allocation to meet needs, relative to the costs of the programme. It also provided a transparent way to weigh these and other factors to make an overall judgement of VfM.

It would have been challenging to evaluate VfM of the SNG Programme using cost-benefit analysis alone. For example, a true cost-benefit analysis would require valuation of the impacts of health and education services on the lives of citizens (Drummond et al., 2005). These benefits will be realised beyond the term of the programme, will be problematic to attribute, and some impacts may be difficult to value monetarily.

The process of engagement about values—determining what is important to judge the performance of the programme—did more

than provide a robust conceptual framework; it also provided the opportunity for open conversations about the programme that built relational trust between the evaluators and the programme teams, encouraging candid discussion and honest reflection.

Despite the intended long-term outcome, that “*poor people in Punjab and Khyber Pakhtunkhwa report that services are better meeting their needs*”, the VfM evaluation did not involve direct engagement with citizens. This reflects the nature of the programme, which focuses on reforms in public financial management, governance, and planning. The intermediary outcomes which were the primary focus of this VfM assessment will largely be invisible to citizens, and their downstream effects on service delivery, and public satisfaction with services, will be evaluated separately. Nevertheless, citizen engagement will become increasingly important if further VfM assessments of SNG are carried out in the longer term.

When generalising this approach to other programmes and contexts, consideration must be given to proportionality. The use of explicit evaluative reasoning to guide evaluative judgements can be tailored to available resources for conducting VfM assessments; the underlying reasoning process remains the same while the applicable criteria and the content and comprehensiveness of evidence can be varied.

The use of rubrics may be new to VfM assessment in aid programmes, but is an example of explicit evaluative reasoning, which is core to good evaluation (Yarbrough et al., 2011). When VfM is assessed in this way, there is a greater overlap with general M&E. Embedding VfM assessment with the programme’s wider M&E processes can ensure the VfM and other M&E frameworks cohere as well as promoting efficient data collection processes. Although much of the evidence needed for the VfM assessment drew on existing programme data and documentation, the collation of data into the

necessary format required considerable time and effort for the 2017 VfM assessment. In the 2018 assessment, however, VfM reporting was better harmonised with other routine reporting and was able to be undertaken more efficiently.

Ultimately, evaluation is only of value to the extent that it is useful and used. By making VfM judgements transparent, the application of explicit evaluative reasoning in the SNG Programme and in other programmes (e.g., King & Guimaraes, 2016) has been well received by DFID annual review teams and underpinned the evidence used by DFID to assess performance. The approach provides a structure for evaluative reasoning that is sufficiently flexible to be applied to diverse complex systems, where the objective is to assess the value derived from investments where the relationship between cause and effect is non-linear.

Acknowledgements

The authors would like to acknowledge the UK Department for International Development (DFID), Oxford Policy Management (OPM) and the SNG teams of Punjab and Khyber Pakhtunkhwa for their participation in the VfM assessments and for permission to share this account. Professor John Hattie, University of Melbourne, and Matthew Powell, Senior Consultant in OPM's Governance Programme, reviewed drafts and provided helpful feedback. Any errors or omissions are the responsibility of the authors.

References

- Adler, M. D., & Posner, E. A. (2006). *New foundations of cost-benefit analysis*. Cambridge, MA: Harvard University Press.
- Adou, J. Y. (2016). Applications and limitations of value for money in development activities. *eVALUation Matters, Third Quarter*, 6–21. Africa Development Bank.

- Andrews, M. (2013). *The limits of institutional reform in development: Changing rules for realistic solutions*. Cambridge: Cambridge University Press.
- Archibald, T., Sharrock, G., Buckley, J., & Young, S. (2018). Every practitioner a “knowledge worker”: Promoting evaluative thinking to enhance learning and adaptive management in international development. In A. T. Vo & T. Archibald (Eds.), *Evaluative thinking. New Directions for Evaluation*, 158, 73–91.
- Bayes, T., & Price, Mr. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London*, 53(0), 370–418.
- Befani, B., & Mayne, J. (2014). Process tracing and contribution analysis: A combined approach to generative causal inference for impact evaluation. *IDS Bulletin*, 45(6).
- Befani, B., & Stedman-Bryce, G. (2017). Process tracing and Bayesian updating for impact evaluation: *Evaluation*, 23(1), 1–19.
- Davidson, E. J. (2005). *Evaluation methodology basics: The nuts and bolts of sound evaluation*. Thousand Oaks, CA: Sage.
- Davidson, E. J. (2014). Evaluative reasoning. *Methodological Briefs: Impact evaluation 4*. Florence: UNICEF Office of Research.
- Department for International Development (DFID). (2011). *DFID’s approach to value for money (VfM)*. London: Author.
- Department for International Development (DFID) Pakistan. (2012). *Sub-national Governance Programme: Business case*. Islamabad: Author. Retrieved from <https://devtracker.dfid.gov.uk/projects/GB-1-202367/documents>
- Drummond, M. F., Sculpher, M. J., Torrance, G. W., O’Brien, B. J., & Stoddard, G. L. (2005). *Methods for economic evaluation of health care programs*. Oxford: Oxford University Press.
- Fleming, F. (2013). Evaluation methods for assessing value for money. Accessed from <http://www.betterevaluation.org/en/resource/assessing-value-for-money>

- Fournier, D. M. (1995). Establishing evaluative conclusions: A distinction between general and working logic. In D. M. Fournier (Ed.), *Reasoning in evaluation: Inferential links and leaps. New Directions for Evaluation*, 58, 15–32.
- Independent Commission for Aid Impact (ICAI). (2011, November). *ICAI's approach to effectiveness and value for money*. London: Author.
- Independent Commission for Aid Impact (ICAI). (2018, February). *DFID's approach to value for money in programme and portfolio management: A performance review*. London: Author.
- Julnes, G. (2012). Promoting valuation in the public interest. In G. Julnes (Ed.), *Promoting valuation in the public interest: Informing policies for judging value in evaluation. New Directions for Evaluation*, 133, 109–129.
- King, J. (2017). Using economic methods evaluatively. *American Journal of Evaluation*, 38(1), 101–113.
- King, J., & Guimaraes, L. (2016). Evaluating value for money in international development: The Ligada Female Economic Empowerment Programme. *eVALUation Matters*, Third Quarter, 2016, 58–69. Africa Development Bank.
- King, J., McKegg, K., Oakden, J., & Wehipeihana, N. (2013). Rubrics: A method for surfacing values and improving the credibility of evaluation. *Journal of MultiDisciplinary Evaluation*, 9(21), 11–20.
- King, J., & OPM VfM Working Group. (2018). OPM's approach to assessing VfM: A guide. Oxford: Oxford Policy Management Ltd. Available at www.opml.co.uk/publications/opm's-approach-assessing-value-money
- Mayne, J. (2008). Contribution analysis: An approach to exploring cause and effect, *ILAC methodological brief*. Available at www.cgjar-ilac.org/files/ILAC_Brief16_Contribution_Analysis_0.pdf
- Nicholls, J., Lawlor, E., Neitzert, E., & Goodspeed, T. (2012, January). *A guide to social return on investment*. Haddington, UK: The SROI Network.
- Rogers, P. (2014). Theory of change. *Methodological Briefs: Impact Evaluation 2*. Florence: UNICEF Office of Research.

- Schiere, R. (2016). What is new in value for money? *eVALUation Matters*, Third Quarter, 2016, 22–33. Africa Development Bank.
- Schwandt, T. (2015). *Evaluation foundations revisited: Cultivating a life of the mind for practice*. Redwood City: Stanford University Press.
- Schwandt, T. A. (2018). Evaluative thinking as a collaborative social practice: The case of boundary judgment making. In A. T. Vo & T. Archibald (Eds.), *Evaluative thinking. New Directions for Evaluation*, 158, 125–137.
- Scriven, M. (1991). *Evaluation thesaurus*. Newbury Park, CA: Sage.
- Scriven, M. (1995). The logic of evaluation and evaluation practice. *New Directions For Evaluation*, 68(Winter), 49–70.
- Van Evera, S. (1997). *Guide to methods for students of political science*. Ithaca NY: Cornell University Press.
- Vo, A. T., Schreiber, J. S., & Martin, A. (2018). Toward a conceptual understanding of evaluative thinking. In A. T. Vo & T. Archibald (Eds.), *Evaluative thinking. New Directions for Evaluation*, 158, 29–47.
- Yarborough, D. B., Shulha, L. M., Hopson, R. K., & Caruthers, F. A. (2011). *The programme evaluation standards: A guide for evaluators and evaluation users* (3rd ed.). Thousand Oaks, CA: Sage.

The authors

Julian King, Director, Julian King & Associates Ltd—a member of the Kinnect Group; Associate, Oxford Policy Management Ltd; Honorary Fellow, The University of Melbourne.

Email: jk@julianking.co.nz

Stephanie Allan, Senior Consultant, Public Financial Management, Oxford Policy Management Ltd.

Email: Stephanie.allan@opml.co.uk