Briefing Note

Indicators and VFM in Governance Programming

July 2011

INTRODUCTION

This note provides advice to DFID staff on:

i. governance indicator best practice, and
ii. measuring the Value for Money of governance programmes

This note is for use primarily by DFID governance advisers, as well as other DFID staff designing programmes with governance elements. The note provides a framework for consideration in Business Case design that relates to governance activity.

On Value for Money (VFM) in particular, this guidance is only intended as 'interim' whilst further research is undertaken. During 2011-2012, DFID will work to determine best practice and establish agreed approaches and mechanisms. This guidance will therefore be updated accordingly subject to research findings as they are made available.

This note was drawn up by DFID staff. It builds on 2 research reports by ITAD, submitted in December 2010 and January 2011 respectively,[1] as well as DFID's internal Business Case guidance. There are 2 main sections: **Section 1: Governance Indicators** and **Section 2: Value for Money in Governance Programming**. The note ends with 10 Top Tips on Business Case preparation.

One of the most frequent requests for support on indicators and VFM is simply to be told 'how to do it'. Given the context-specific nature of most DFID activity, and in particular governance programming support within that, it is frequently not possible to provide guidance that simply sets out a 'How To' set of instructions. Instead, this note should be viewed as providing key points to consider and take into account during programme design. Points should then be adapted to specific programme design.

If you work in fragile and conflict-affected countries, you may also want to refer to guidance on Results in fragile and conflict-affected states. This sets out good practice in ***all*** programming – not just governance - on a) measuring the impact of programmes on a country's conflict and fragility and b) measuring results in contexts that are difficult in terms of security, logistics and capacity. The guidance also includes pointers on VFM for conflict and fragility programming.

For any questions, comments and/or revisions to this note, please contact the Governance Gender and Fragility **Results Hub**, in the Politics, State & Society Team (PSST), in Policy Division.

[1] Barnett, C; Barr, J; Christie, A; Duff, B; Hext, S (2011), "*Measuring the Impact and Value for Money of Governance and Conflict Programmes*", DFID (Quest doc: 3120325); and Barnett, C; Barr, J; Christie, A; Duff, B; Hext, S (2011), "*Governance and Conflict Indicators Report*", DFID (Quest doc: 3120333)

CONTENTS

ACRONYMS

3 Es	Economy – Efficiency – Effectiveness (the UK NAO's framework for VFM)
CBA	Cost-Benefit Analysis
CPI	Transparency International's Corruption Perceptions Index
CSO	Civil Society Organisation
CSR	Civil Service Reform
DFID	UK Department for International Development
DRC	The Democratic Republic of the Congo
E&A	Empowerment and Accountability
M&E	Monitoring and Evaluation
NAO	National Audit Office
NGO	Non-Governmental Organisation
PEFA	Public Expenditure and Financial Accountability
PETS	Public Expenditure Tracking Survey
PFM	Public Financial Management
PRSP	Poverty Reduction Strategy Paper
PSST	Politics, State & Society Team
RBM	Results-Based Management
S&J	Security and Justice
UPR	Universal Periodic Review (Human Rights)
VFM	Value for Money
WGI	Worldwide Governance Indicators

SUMMARY

1. **Global governance indicator datasets are weak**; in turn DFID reporting systems do not adequately capture and record the impact of our governance programming. We are working to address both these deficiencies.

2. **A good set of programme-level governance indicators** has both specific and general characteristics. Indicators will be: clear; rule-bound; causally linked; gender and pro-poor in focus. They should also take political assumptions; risk; sectoral linkages; and participation into account.

3. **Common errors in indicator design** include conflating indicators and objectives; confusing output and outcome indicators; and overly-ambitious outcome setting. These errors can be avoided if general indicator guidance is taken into account, including ensuring programmes are based on robust analysis, and set out clearly what they want to achieve, and how, i.e. that they are explicit about the proposed theory of change.

4. **Value for Money can be assessed in different ways**. Whilst research and international dialogue are ongoing to agree processes for measuring the results and VFM of governance programming, DFID will use the UK National Audit Office (NAO) '3 Es' approach: this assesses each governance programme against measures of Economy, Efficiency, and Effectiveness.

5. **Good VFM achieves a balance across the 3 Es**: spending less; spending well; spending wisely. It is not about just using the cheapest option. A robust VFM assessment relies on strong data and evidence. In the event such data is not available, mechanisms to find it should be provided for during programme implementation, and related costs factored into programme design.

6. **Measuring VFM: is not new**; takes time and resources; does not preclude risk-taking; and should be proportionate. It is about using a common sense approach to programme design. DFID is not asking 'What's the VFM of an election?' Instead, DFID is asking that any support to an election – in whichever country – uses UK funds as effectively as possible, based on credible data.

7. **An analytical framework for VFM** is made up of several elements: the key components of a programme; VFM measures; modifiers for context and assumption; DFID's contributor share; and confidence levels in the data. The framework recognises that well-designed governance programmes are based on robust social, economic and political analysis, and that DFID does not act alone when supporting reform. It also allows DFID to begin to assess impact even when a situation has deteriorated, i.e. to set out how a situation may have been worse without DFID's intervention.

8. **Four, cumulative, 'Options for VFM'** are presented: improved results-based management (RBM); a ratings and weightings approach; trends analysis; and cost benefit analysis. Ideally, programme designers should aim to conduct a cost-benefit analysis and as a rule of thumb, only the absence of sufficient and robust data should mean a move to a less robust VFM assessment. Research is ongoing to assess the most appropriate VFM methodologies across the various 'sub-themes' of governance.

SECTION 1: GOVERNANCE INDICATORS

1.1 Background

9. ***Useful but limited global indicator datasets; weak internal recording systems***. Internationally, the term 'governance indicator' is used in different ways. It can refer to international datasets on governance; a combination (or 'basket') of individual indicators to measure specific programme outputs and/or outcomes; or the individual indicator itself. This report uses the term 'governance indicator' to mean individual indicators of activity at the programme level of DFID activity.

10. Until recently, DFID relied on a number of international datasets to monitor governance trends in its partner countries, and analyse political context. Such datasets include the Worldwide Governance Indicators (WGI), housed at the World Bank Institute; Freedom House's 'Freedom in World' scores; Transparency International's Corruption Perceptions Index (CPI) and Global Corruption Barometer; the Bertelsmann Transformation Index; and others. Each of these datasets has methodological flaws, but nonetheless, provides useful information about the direction of travel on a variety of issues in DFID partner countries.

11. What they cannot do, however, is indicate why. They cannot point out, for example, *why* perceptions of corruption have gone up or down, or *why* a country's political stability has improved / deteriorated. They are not designed to. So, when it comes to monitoring DFID's impact through a particular programme (or that of any other actor), these datasets are of limited use: they are set at too high a level to provide *attributable* impact information.

12. To date, this has been acceptable: it was recognised that measuring governance trends is notoriously difficult (hence the methodological flaws in all existing governance datasets) and that attempting to aggregate results across the broad expanse of governance activity risked under-emphasising specific results; missing the point of tailoring intervention to context; and stretching out global commonalities to the point of banality. DFID reporting on governance therefore remained unattributable, and based on international datasets.

13. Such practice in DFID, however, also led to an under-reporting of *actual* impact. Recording what we did and achieved through governance programming was not mainstreamed into annual and project completion reports. Over time, this meant that DFID was unable to report meaningfully on its governance activity, whilst at the same increasing its governance spend. In accountability terms, this was not sustainable.

14. As a first step to address this situation, logframes from a range of governance programmes were examined to see which indicators were being used in practice. Based on this, as well as taking into account existing indicators from other organisations, lists of suggested indicators for use in DFID programmes were drawn up across the range of governance sub-themes (Security & Justice [S&J]; Civil Service Reform [CSR]; Empowerment & Accountability [E&A]; etc). These suggestions were then tested for relevance and robustness. A revised list is attached at Annex 1.

15. The findings of the testing of the suggested indicator lists are set out below. Note that suggested indicators remain just suggestions. They have been specifically drawn up to be adjusted according to both context and individual programme needs.

1.2 Indicator Testing – Findings

16. In being tested for robustness and relevance, it was found that sets of governance indicators within a particular programme should have a number of characteristics, both specific and general.

17. **Specific indicator characteristics**. Indicators should be:

    - ***Clear***
    - ***Rule-bound***
    - ***Causally-linked***
    - ***Gender and/or pro-poor in focus***

18. ***Clear*** indicators are precise units of measurement that do not set direction. They specify what is to be measured, rather than what is to be achieved. Some observers use the terminology 'specific' and 'measurable' when discussing this aspect of indicators. In Table 1 below we see examples at outcome and output level of clear and unclear indicators. The better indicators are clearer: it is obvious what is to be measured and monitored over time. There is no sense from the 'unclear indicators' of *how* the outcome or output can and/or will be measured.

Table 1: Clear and Unclear indicators

	Clear indicator	Unclear indicator
Outcome	% people who feel safe going out at night	Extent of respect of code of conduct by political parties
Output	# reported incidences of violence	Quality of training provided

19. Note that by being clear, or measurable, indicators often become quantifiable. This is important for later stages of programme management, for example when beginning to assess the VFM of programmes (in both design and evaluation stages). It is also important in keeping ambitions and statements about intended outcomes realistic. It is likely that by setting out clear, measurable indicators of programme impact, DFID will see a natural lowering of ambition in outcome statements.

20. ***Rule-bound***. Indicators should measure at the correct level: output-level indicators should measure outputs and outcome-level indicators should measure outcomes. Whilst this seems an obvious thing to say, it is perhaps one of the hardest elements of programme design: DFID experience shows there is much confusion between output and outcome indicators. This partly reflects the context specificity of each programme – an output in one programme may well constitute an outcome in another programme – but equally, it reflects that the logic within governance programming design has not always been as robust as it might have been. For an example, see Table 2.

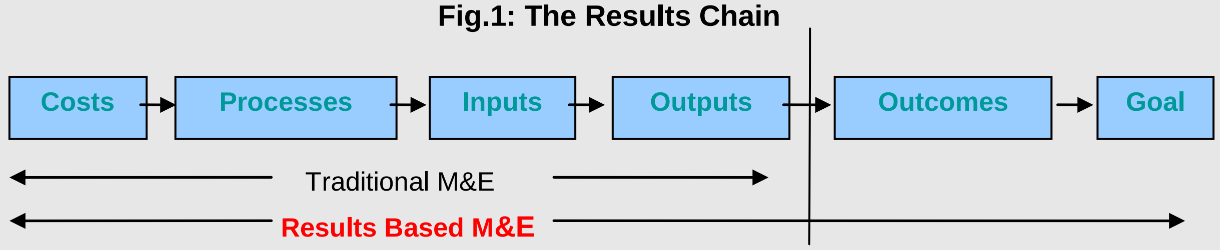
Table 2: Indicators should measure the appropriate level

Outcome	% citizens expressing trust in formal rule of law institutions
Output	# cases resolved by provincial courts

21. In this case, it would not be appropriate to use '# cases resolved' at the outcome level unless your programme intended impacts were very unambitious! Equally, an objective of 'increased trust' would not be appropriate at output level as individual activities are unlikely to be able to achieve such a direct, qualitative change.

22. This example also serves as a useful reminder that all elements of a results chain should set out the ***theory of change*** that sits behind a programme of support – and that all elements require indicators. The unwieldy aspect of 'trust' is being measured *clearly* (as per paragraph 18) – through an indicator that can be measured quantitatively. Whilst the programme *objective* may be 'improved level of trust' or 'better rule of law institution', the *indicator* is not expressed in this way. An indicator stated as 'quality of rule of law institution' or 'level of trust', for example, would not be clear – neither of these would usefully show *how* the outcome is to be measured. So, instead, the indicator of (in this case) trust becomes '% citizens expressing trust'. It is not a perfect indicator, but results will usefully show how trends may change over time. The example serves as a reminder that the qualitative nature of much governance work can be measured, even at outcome level.

Fig.1: The Results Chain



23. ***Causally-linked***. Indicators should also link to each other up the results chain: just as in programme design of whatever kind inputs should lead to processes (sometimes called activities), which in turn should lead to outputs, which lead to outcomes, which lead to overall goals, so it is with respective indicators. Output indicators should link to outcome indicators and make the programme's theory of change explicit.

24. In Table 3 below, for example, we see that DFID is supporting a programme that anticipates more Parliamentary seats for women. The programme's theory of change is implied: (an increase or decrease in) the number of seats held by women is linked to the number of women candidates standing for election.

Table 3: Indicators should reflect a programme's theory of change

Outcome	% seats in parliament held by women
Output	# women candidates in an election

25. We can also use this example to highlight that programme theories of change(s) may be flawed: it may be that after programming has started that it is shown – for whatever reason - that the number of women candidates has no bearing on the % of seats in parliament held by women. In this case, either the programme should be redesigned (support women candidates in a different way), or the underlying theory of change needs testing and revisiting (stop supporting women candidates). Either way, using indicators of this kind to monitor a programme's progress will assist in general programme management.

26. ***Gender and pro-poor focus***. Testing of the draft suggested lists of governance programme indicators also showed that DFID programmes could make a focus on gender and pro-poor outcomes much more explicit than they have in the past. We know that overly-technical approaches to any aspect of development, as against acknowledging underlying power dynamics, are unlikely to bring about desired outcomes. Indicators should measure change in institutional structures that underpin

the distribution of power and shape relations between state and societal actors, including those that might empower women, the poor and/or other excluded groups.

27. For example, instead of designing a programme that uses measurable outcome indicators, and then disaggregating the results by sex, a programme could more expressly address power inequality by measuring institutional changes. See Table 4:

Table 4: Indicators should measure institutional change from a gender and/or prop-poor perspective

Outcome	% seats in parliament held by women
Output	# criminal cases of domestic violence

28. Note that, in this case, it is very unlikely that only work on criminal cases on domestic violence will lead to a shift in parliamentary seats – instead, work on domestic violence may be only one element of a broader programme! This is discussed in more detail in paragraph 38 below. The point to take away here is simply that DFID's earlier logframes could have been more specific about the gender (and/or pro-poor) focus of the intervention.

29. **General indicator characteristics**. In addition to the specific characteristics of indicators set out above, there are some general characteristics of indicators that need to be considered during design. These are:

   - ***Political assumption and risk***
   - ***Participation****; and*
   - ***Cross-sector linkages***

30. ***Political Assumption and Risk***. Theories of change, however narrowly defined, are still contingent on a range of assumptions holding true. Indicators that take political assumptions into account (i.e. what is more or most likely to happen in a particular context) will likely engage with any perverse incentives that could come into play. This will contribute to better risk management - if you are aware of context, you are aware of potential risks, and can mitigate against them.

31. For an example, see Table 5. In this case, the indicators reflect a programme theory that neutral journalists play a role in improving government accountability.

Table 5: Indicators should take political assumption into account

Outcome	% citizens expressing improvement in government accountability
Output	# journalists aware of their responsibilities for neutrality

32. There are several implied assumptions here (that would need to be set out in programme documentation) such as 'journalists can operate freely in the environment'; 'journalists that are aware of neutrality responsibilities act on them'; etc. Once set out, programme designers can reflect if assumptions are appropriate or realistic in a given context and mitigate against any related risks as necessary.

33. We can use this same example to illustrate a linked but separate point – that aggregating results from governance indicators can be difficult. It is not correct to say that aggregation is never possible or appropriate, but nonetheless, the context specificity of governance programming makes results aggregation harder than in other areas of programming. There are two broad reasons for this: disparity of reporting 'level', and different indicators being more appropriate to measure similar outcomes at the same reporting level.

34. So – on the disparity of reporting level point, we've seen above (paragraph 20) that outputs in one programme may constitute outcomes in other programming, so that results measured across the board at one 'level' of reporting (for example, all outcome indicators on accountability) may miss impacts achieved elsewhere, in this case at output level on accountability in other programmes.

35. And, in terms of different indicators being appropriate in different contexts, it may be that, to use the example in Table 5, the indicator '% citizens expressing improvement in government accountability' is simply not the best measure of improved government accountability elsewhere. A more appropriate indicator may be, say, '# of audit reports published and accessed'; or '% parliamentarians who have complied with asset declaration requirements' – i.e. whichever measure is appropriate to the programme and context. This means that any aggregating all results of '% citizens expressing improvement in government accountability' will simply not reflect all impact of various government accountability programmes.

36. In itself, this is OK – or at least for now. DFID will be conducting research into this tension during 2011 and 2012, to enable improved reporting of governance programme impact, as well as to learn about how best to aggregate results.

37. ***Participation***. Change is often a matter of perspective. Who decides what to measure is therefore important. It stands to reason that those who have a stake in the change should be consulted on what the change looks like. This is basic good practice in any programme management, so – where possible – stakeholders should be included in programme design and, ideally, set out the indicators to be used to monitor progress. Where this is not possible, stakeholders should at least be aware of how progress is to be monitored.

38. ***Cross-sector linkages***. Indicators can be used across either various governance sub-themes, or across the remit of a DFID portfolio in country. I.e. it is likely that more than one output leads to a broader outcome, and a 'basket' or selection of indicators should be used to set this out. 2 examples to illustrate this are:

- ***Within governance programming***, a theory of change on women's representation in parliament is unlikely to be justified *only* through a change (increase or decrease) in the number of domestic violence cases. More likely, according to context, a programme that looks to impact on the proportion of parliamentary seats held by women may well be justified against a theory of change that combines domestic violence levels (S&J programming), numbers of women candidates supported in an election (elections programming), and numbers of civil society organisation (CSO) published articles advocating women's rights (E&A programming). See Table 6.

Table 6: A basket of indicators can cut across sub-themes of governance to demonstrate robust monitoring of programme outcomes

Outcome	% seats in parliament held by women
Output	# women candidates in an election # CSO articles advocating women's rights # criminal cases of domestic violence

- ***Across DFID programming***, a theory of change on improved education service delivery may draw on governance-specific and sectoral outputs. For example - again, according to context - it may be that a programme to increase the number of children completing primary school is justified against a results chain that measures the proportion of children who can read; the numbers of teachers trained (education outputs) and the proportion of education funding reaching targeted schools (governance output). See Table 7.

Table 7: A basket of indicators can cut across DFID's portfolio of programming to demonstrate robust monitoring of programme outcomes

Outcome	# children completing primary school
Output	% children who can read simple text after 2 years in school # teachers trained % education funding reaching targeted schools

39. This logic can be applied at any stage of the results chain. So, just as in the examples above, baskets of output indicators are used to achieve outcomes, a basket of outcome indicators can be used to achieve wider 'goals'. Building on the examples in Tables 6 and 7, it could be argued that the goal of 'improved government effectiveness' (measured by using a country's WGI ranking as an indicator) is achieved through improved education service delivery and increased women's representation in parliament. See Table 8.

Table 8: Just as baskets of output indicators can contribute to outcome monitoring, so can baskets of outcome indicators contribute to goals

Goal	WGI 'Government Effectiveness' score
Outcome	# children completing primary school % parliamentary seats held by women

40. Again, for programming purposes, whatever the context, the key point to remember is that programme logic is sound; presented clearly; and is applied consistently across the relevant range of programmes / business cases.

1.3 Common Errors in Indicator Design

41. The testing of the draft suggested governance indicators showed a tendency in DFID towards 3 common errors:

- ***Conflating indicators with objectives***
- ***Confusing output and outcome indicators***
- ***Overly-ambitious outcome setting*** (e.g. 'gender equality achieved')

42. It is suggested that these errors are linked. Indicators that are unclear, or overly-ambitious will likely use ambiguous terms, such as 'level of' or 'access to'. They suggest a need to revisit programme objectives, and a fuller consideration of the programme theory of change.

43. We know already from the results chain that outputs should be causally-linked to outcomes: if output A happens, then outcome B will be achieved. Indicators reflect the causal link but they are not the same as the causal link – an indicator shows how A and B will be measured (so cannot itself be 'A' or 'B'). **If objectives are being stated as indicators, it is likely that objectives require greater specification**.

44. It is likely that by improving practice on indicators, DFID will see a natural lowering of ambition in outcome statements. This is welcome: over time, it will mean that governance programming shifts from largely being based on theory, to being more substantively based on evidence.

1.4 Suggested indicators

45. A revised list of suggested indicators for governance programming is presented at Annex 1. DFID expressly recognises that there are a number of issues to consider with caution when producing this type of list, for example: that only suggested indicators are used in programming; or that programme focus reduces to a standardised set of limited outcomes; or even that political context and analysis are not taken into account in programme design. It is hoped that by following the main points highlighted above, these concerns can be allayed.

46. Finally, further work is still needed, to be done in slower time, to address the remaining flaw in current (global) governance indicator practice, i.e. understanding how a result from a DFID (or other actor) programme impacts on country-level changes, as set out in international datasets, like the CPI or WGI. These questions will be addressed by DFID Policy and Research & Evidence Divisions during 2011 and 2012.

Indicators Summary

When choosing how to monitor programme progress, a good set of indicators will have specific and general characteristics. *Specifically*, they should be

- **Clear** (what is being measured?)
- **Rule-bound** (is the indicator appropriate to level of monitoring?)
- **Causally-linked** (is the programme theory of change set out explicitly?) and
- **Gender and pro-poor** in focus (don't just disaggregate by sex!)

Indicators should take political context into account, and be appropriate for the environment. And, ideally, they should be chosen or designed with relevant stakeholders involved. *Generally*, a good set of indicators will be

- **Politically realistic** (will this work here? Has analysis been taken into account?)
- **Participative** (have relevant stakeholders helped design these indicators?)
- **Cross-sectoral** (how does this programme fit with broader objectives?)

SECTION 2: VALUE FOR MONEY IN GOVERNANCE PROGRAMMING

47. There are various different ways to measure Value for Money. It is not an exact science, and an element of judgement will always be involved. The key message from DFID's Chief Economist is to take a common sense approach to this aspect of programming: show you are using all available evidence and have considered VFM, and programme review will progress smoothly.

48. This section sets out a broad framework with which to approach measuring the VFM of governance programmes, and moves on to describe various methodologies for measuring VFM. Much of the content below is applicable across any type of programming, and is not unique to governance.

2.1 What is VFM?

49. VFM is a set of assessment practices for appraisal, review or evaluation of systems and functions. It is closely linked to results-based programme management. The UK government uses VFM to reflect 2 concerns:

    - Increased transparency and accountability in spending public funds; and
    - Obtaining the maximum benefit from available resources

50. The UK's National Audit Office (NAO) defines VFM as being 'the optimal use of resources to achieve intended outcomes'. In turn, DFID has defined VFM as:

    "*maximising the impact of UK aid so it makes the most difference to the poorest people in the world*".[2]

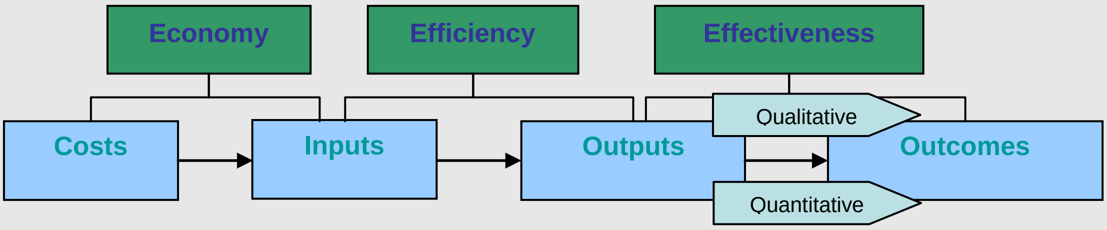
51. Early VFM work focused on economy savings, but this is not true today. DFID is clear that measuring VFM is not about a 'race to the bottom' and using the cheapest option in programme implementation. The question 'What's the VFM of an election?' is not being posed specifically. Nor is support to an election in, say, Ghana, being prioritised over support to an election in DRC because operating costs in Ghana may be lower. And nor is support for more expensive programming 'types' being deprioritised – for example elections support being deprioritised against Public Financial Management (PFM) reform, given the (generally) lower operating costs in PFM programmes.

52. Instead, by renewing its focus on VFM, DFID is asking that we get better at understanding – and thereby continually reviewing - our costs: once the evidence (based on analysis) has been presented that an elections support programme is appropriate, programme documentation needs to show that any such support – in whichever country – ensures that UK funds are used as effectively as possible, based on credible data. ***Demonstrating VFM is about (improving) aid effectiveness***.

53. The UK National Audit Office (NAO) uses a 3-pronged approach to VFM, sometimes known as the '3 Es', which examines not just economic factors, but also efficiency and effectiveness. ***VFM is about spending less (economy); spending well (efficiency); and spending wisely (effectiveness)***. Putting this onto the results chain looks like this:

[2] DFID Value for Money Core Script, March 2011

Fig. 2: VFM assessment on the Results Chain



54. A good VFM assessment achieves a balance across the '3 Es': it is not the case that the cheapest option is always the best VFM. Instead, "*VFM is high when there is an optimum balance between all three elements, when costs of relatively low, productivity is high, and successful outcomes have been achieved*".[3] The NAO states "*good value for money is the optimal use of resources to achieve the intended outcome*."[4]

55. This means that VFM decisions necessarily involve judgement – judgement based on evidence. Essentially, this is good programme management practice, and is not particularly new: if you can clearly set out why you plan to programme UK funds, how they will be used, and the intended impact, you will have the evidence you need to set out why such a programme constitutes good VFM.

2.2 How will DFID use VFM?

56. VFM assessments are generally used in 2 ways:

- o Programme Appraisal
- o Evaluation of programme performance

DFID will use both – first as part of Business Case preparation, to decide whether to invest in a project, and later, as part of normal project cycle management to evaluate impact. The two are linked: **unless VFM consideration is made during programme design, it will not be possible to assess the programme for VFM later, during evaluation**.

57. A programme's design should therefore set out the criteria and evidence against which the programme will later be judged. The overall VFM judgement is arrived at by comparing actual performance against the planned / anticipated outcomes. The quality of a VFM assessment depends on the quality of data put in. This is the principal difference between measuring the VFM of governance, as opposed to any other kind of, programming: evidence on governance activity is usually weak, and may be contested, due to the contextual specificity required in all programming. It is easier to use comparative data in a health or education programme both because some elements of such programming are universal, and because data have been being produced for decades. In the case of governance programmes, cost and outcome comparisons have usually not yet been made, so establishing the VFM of governance support programmes will require considerable effort going forwards.

[3] Audit Commission Webpage, accessed 22 February 2011
[4] NAO VFM Analytical Framework

58. ***The link to indicators***. There is a significant role for indicators in VFM assessment. Indicators provide both measures of productivity (efficiency) and qualitative and quantitative measures of outcomes (effectiveness). If a DFID programme is designed to use indicators that are unclear, or otherwise of low quality, it will not be possible to evaluate that programme for impact and VFM later on.

59. ***Proportionality***. Getting to a 'good' VFM assessment requires

   - ***robust contextual analysis***
   - ***good quality data*** – of the country; previous experience of related sector programming; and up-to-date research
   - ***time***.
   - ***funds***. It is not credible to undertake a VFM assessment as part of programme evaluation without devoting adequate resources to fund the process. This should be factored into programme design.

60. Any approach to measuring the VFM of either a proposed or completed programme should not be 'crowded out' by a fixation on numbers and quantitative assessment. Measuring VFM should also <u>not</u> be viewed as an 'add on', to be considered quickly, once programme design is complete.

61. ***It is possible to spend a disproportionate amount of time on VFM assessments***. Clearly, large-scale expenditure of UK funds requires good quality consideration of VFM, and appropriate levels of staff time and input should be used accordingly. For smaller-scale programmes, less time should be allocated for VFM assessment. There is no right or wrong here, and as DFID gets more familiar with business case preparation, further knowledge on this area should become apparent.

62. Justification for reduced VFM measurement should be provided: again, ***use a common sense approach***. As long as the reasoning behind any decision on VFM is clearly set out in programme documentation (be it how an assessment is reached, or how much time was allocated to VFM measurement), then programme scrutiny will progress smoothly.

63. ***Risk***. Measuring the VFM of any proposed programme should not put people off acting quickly or taking risks. Clearly, if DFID begins to reduce its appetite for risk, then we potentially miss out on programming activity that could be genuinely transformative. This is particularly relevant for programmes in fragile or conflict-affected countries, where pressures to move fast and to be seen to be 'doing something' are often higher than elsewhere (and therefore where the incentive to 'skip' the VFM measurement of programming will be higher). Calculating the VFM of a proposed programme can help factor in risk – by considering all elements of programme design (what type of intervention; how to implement; levels of expenditure etc) you are factoring in risk.

64. Clearly, though, it is not possible to factor in all risks to programme design. At the end of the day, all VFM assessments entail some numbers, and some judgement - based on the data provided. For business case preparation purposes (and subsequent programme evaluation), **what needs to be ensured is that the reasons for proposed programming decisions are clearly set out**. One way to approach this is by working through a framework for VFM. See next section.

2.3 Analytical Framework for VFM in Governance Programming

65. The framework presented below is one reading of the NAO's 3 Es approach to measuring VFM. More work is needed to know how best to measure the VFM in the various sub-themes of governance activity. This will be researched in coming years. For now, **this framework should be considered a starting point**. As research becomes available over time, and as international dialogue on this topic continues, this framework may evolve and will be updated as necessary.

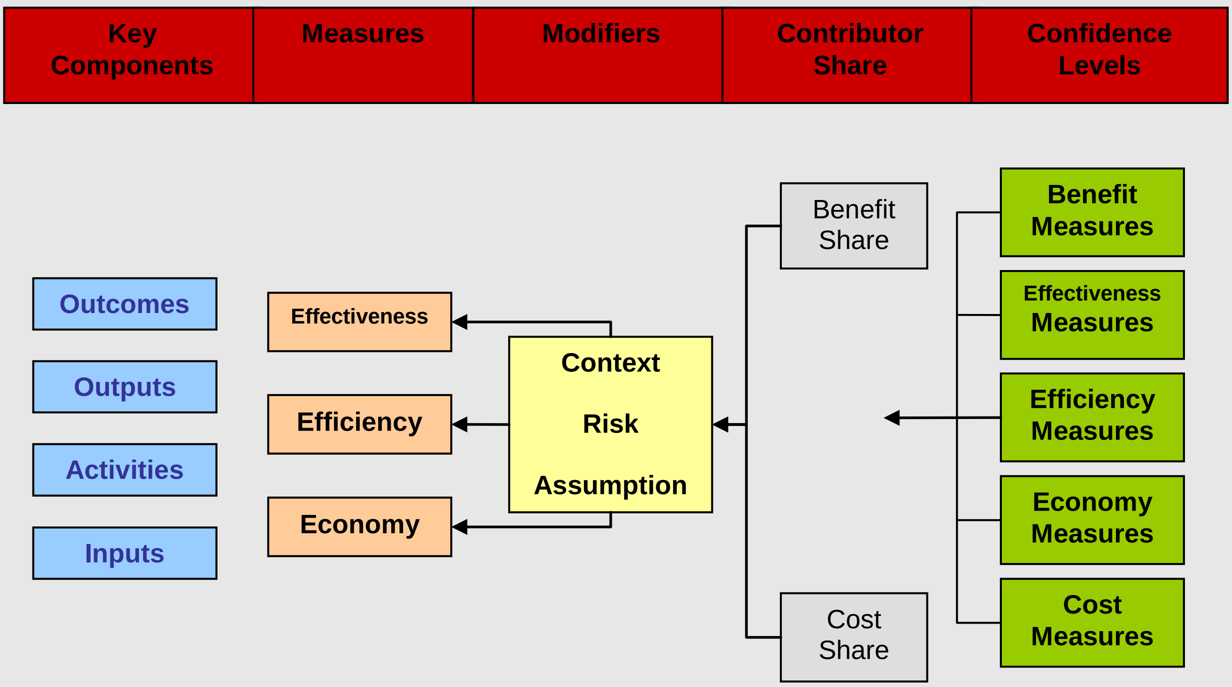
66. The framework incorporates 5 elements. The first 2 correspond to the general VFM measurement framework (discussed above in Figure 2), i.e.

- o **Key components** of any programme (as set out in the results chain: inputs; activities; outputs; outcomes); and
- o **Measures** of VFM – the '3 Es' of Economy, Efficiency, Effectiveness

67. Adapting this to governance programming, the framework also incorporates

- o **Modifiers** for context, risk and assumptions, to take into account the inherently political nature of development work: (will this programme work in this particular context?)
- o **Contributor share**. DFID recognises that it does not act alone: taking this into account means that we can begin to calculate a measure of our impact, including where deterioration has occurred and we want to identify how much worse a situation may have been without the DFID intervention; and finally
- o **Confidence levels** should be considered (how good are the data in each of these framework components? How is VFM affected if assumptions change?)

Fig. 3: Analytical Framework for VFM in Governance Programmes



68. If all respective elements of this framework are considered during programme design and set out in business case documentation, a VFM judgement provided as conclusion will be sufficiently robust.

69. The amount and quality of data available for specific programmes will determine how robust any VFM assessment is. In short, those designing programmes should ask themselves:

   - *Is the theory of change clearly set out against a results chain?*
   - *Are indicators provided to measure each step of the results chain?*
   - *Have the 3 Es been considered, and clearly set out?*
   - *What are the costs and benefits of this programme?*
   - *How confident am I in the evidence used to address each of the above questions?*
   - *Based on all of this, what VFM do I judge this programme to represent?*

70. A hypothetical example of how to use the VFM framework is presented in the box below.

The VFM analytical framework in practice

This example is sets out the sorts of questions to ask in a hypothetical programme supporting improved access to justice through judge training.

- **Economy:** How much is each trainer? Is this a standard cost, globally? What is the cost of hiring a training room? Can you compare this to similar training room hire in country? How do total costs compare to judge training in other countries?
- **Efficiency:** How many judges are trained? Is this the maximum? (If not, why not?) How many trainers are needed? Can this be reduced? (If not, why not?) Can a measure of productivity be devised? What is it?
- **Effectiveness:** Does judge training lead to improved access to justice? How will you know? What is your theory of change? (What is your indicator?)
- **Modifiers:** Contextual matters: is it expensive to run training courses in this context? How do you know? Is it risky to be training judges here? Why? How can you measure this? How far is your theory (training leads to improved access to justice) based on evidence; how far is it assumption?
- **Contributor share:** Even if perception surveys show that people feel that the judicial system is fairer, how will you know that that was down to the training? How do we consider issues of contribution and attribution?
- **Confidence levels:** How representative is the sample survey? How reliable are the data?

71. These questions are not exhaustive. They are presented here to show that by beginning to consider each of these aspects of the VFM framework, you will get closer to a robust VFM assessment. The more data you have – and therefore, the more elements of the framework you are able to consider - the more robust your VFM assessment will be.

2.4 Options for VFM

72. Having established an overall framework, we can now move to various options in VFM methodology. In the contexts in which DFID operates, it is frequently unlikely that there will be sufficient data to undertake a full VFM assessment. In such situations, those designing programmes need to consider what is available quite pragmatically.

73. Ideally, a significant quantity of robust data will be available, allowing for full consideration of anticipated *effectiveness*, and ongoing monitoring of impact – preferably using cost-benefit analysis (CBA). Where this is not possible, best case *efficiency* measures can be considered. And, in a worst case, basic unit costs of *economy* can be compared and considered as sufficient for a VFM assessment. Note, though, that as a rule of thumb, only the absence of sufficient and robust data should mean a move to a less robust VFM assessment.

74. Essentially, this can be viewed in terms of 'working down the results chain'. Programme designers should start by attempting to measure *outcome* VFM (effectiveness); if this is not possible due to lack of data, *output* VFM (efficiency) can be calculated; and if that is not possible, *input* costs (economy) can be calculated and verified for VFM.

75. The more data available, the better impact can be measured. In the case data is not available at programme design stage, efforts should be made to ensure they are gathered during programme implementation. This should be built into programme design, and costed accordingly.

76. One way to present this is through **4 options** to assess the VFM of governance programming (See Figure 4):

Fig. 4: Options for VFM

Management → **Measurement**

A **Improved Results-Based Management**	**B** **Ratings & Weightings**	**C** **Trends Analysis**	**D** **Cost-Benefit Analysis**
Use unit costs of procurement and management procedures Includes: contracting; planning; procurement; management; M&E processes	Develop outcome and cost criteria Develop rating scale and weight accordingly Base VFM judgement on R&W score and narrative	Add in trend information to determine outcomes against a 'do nothing' option Develop outcome statements, based on indicators – requires baseline data Develop rating scale against outcome statements Include trends criteria within benefits table	Outcomes quantified to allow comparison, including against costs Involves other stakeholders Financial proxies developed Underlying trend and baseline data included Sensitivity to assumptions a major issue

A. ***Improved Results-Based Management (RBM)***: using unit costs to measure improvements in results based management and other project management processes (e.g. how much is each ballot paper? Or each day of a consultant's time?)
B. ***Rating and Weightings of the 3 Es***: key processes associated with, and measures of, Economy, Efficiency and Effectiveness are identified and used to rate programmes, and weighted to reflect their relative importance in the programme
C. ***Trends Analysis***: a comparison of progress against indicators is used to measure effectiveness, in which baselines, targets and underlying trends are all taken into consideration
D. ***Cost Benefit Analysis***: outcomes are quantified, often through the use of financial proxies, to make a direct comparison between costs and the financial value of benefits

77. There are **two points to note** about these options:

- ***The options are cumulative***: if data are minimal, then only Option A will be possible. With slightly more than management cost data, it is likely that teams will be able to 'rate and weight'. With sufficient data to allow for regular, robust monitoring, DFID will be able to assess trends and evidence its contributor share. And with sufficient planning and availability of data, DFID will be able to (choose if appropriate to) undertake a full cost-benefit analysis (CBA) of governance programmes

- ***Each option has its strengths and limitations***. Note that even a full CBA is sensitive to assumptions, so risks manipulation to achieve desired results

78. The best place to start to undertake a CBA is to work with your team economist. They are likely to ask many of the questions set out above: what are the costs of intended outcomes; what are the benefits; what would happen if DFID did nothing; etc? It will be your team's judgement about quality and availability of data whether undertaking a full CBA is either feasible or proportionate to the size of the programme. Remember that there is unlikely to be one 'correct' way of undertaking assessing VFM so you will need to consider various measures to decide on the most appropriate way ahead. As more business cases are produced, internal practice will improve in this area.

VFM Summary

A suggested **VFM Analytical Framework** sets out the elements to be addressed in making a VFM assessment:

- ***Key Programme Components*** (the results chain)
- ***VFM Measures*** (Economy; Efficiency; Effectiveness)
- ***Modifiers*** (Context; Risk; Assumptions)
- ***Contributor Share*** (of costs and benefits)
- ***Confidence Levels*** (of data quality)

4 (cumulative) **Options for VFM** have been presented:

- ***Improved Results-Based Management***
- ***Ratings & Weightings***
- ***Trends Analysis***
- ***Cost-Benefit Analysis***

OVERALL CONCLUSIONS AND NEXT STEPS

79. **Much of what is written above is not new**: DFID has been designing and implementing governance programmes for some time! That said, governance indicators and VFM are complex areas, on which no one has 'all the answers'. This note has aimed to provide a few pointers, even reminders, to ensure programme practice is up to date and builds on lessons learned.

80. **This note does not claim to be comprehensive**. Further research on both indicators and VFM is currently being undertaken. In particular, DFID is reviewing if specific VFM methodologies are more appropriate for different 'sub-themes' of governance (elections; PFM; E&A etc). DFID will also start to research unit costs within specific areas of governance programming so that data on comparative costs by country can begin to be built up. Further information will issue during the course of 2012.

81. **Work will also be undertaken internationally**. There is growing interest in improving governance indicators – both at programme and country level. DFID will continually review its use of programme level indicators in different areas of governance programming to consider if some indicators are being used more consistently than others, and the potential implications of such practice. DFID will also commission research to better establish the links between programme level outcomes and country level governance indicators. For example, we need to know how a DFID programme has (or has not) impacted on an overall 'voice and accountability' score, or 'political stability' score. We will be sharing any findings with other donors and programme counterparts.

Governance Indicators and VFM in Business Cases

Top 10 Tips

1. **Think early** about your programme design. It will likely take longer than you think. Don't forget to factor in evaluation methodology from the outset.

2. **Keep it simple**. None of the processes outlined above is strictly new. They simply build on normal programme management practice that, largely, DFID has been following for some time. Try not to over-complicate any aspect of programme design. Ask: Will it make sense to an external observer?

3. **Use common sense**. If it feels overly-complicated, it probably is. Assessing VFM is not an exact science and an element of judgement will always be involved.

4. **Involve your economist and statistician!** All members of a programme design team should be involved right from the start of the process. At a minimum, a renewed focus on results and VFM will mean that governance advisers and economists work together more substantively than in the past. Your team economist will be particularly helpful on VFM assessments; your statistician on checking the robustness of indicators.

5. **Do engage**. Many people unaccustomed to thinking about VFM assessment and evaluation from the start of a programme design process will be tempted not to engage on this agenda. The focus on results and VFM is both a current ministerial priority and will also improve our performance. It is an agenda that is likely to stand the test of time: we need to engage.

6. **Be clear**. Set out what you want to achieve, and how you propose to achieve it. At its most simple, that is your theory of change. (If DFID does X, then Y will happen.) With robust indicators alongside to monitor progress, you can assess VFM, evaluate robustly, and contribute to building up the (global) evidence base on governance programming.

7. **Be proportionate**. For larger programmes, it is likely that design and VFM calculation will necessitate more time and resources than smaller-scale ones. Don't spend days on VFM assessments for lower-priority programmes.

8. **Be consistent**. More than one business case may contribute to wider DFID portfolio goals. This is entirely acceptable as long as your individual programme logic and assumptions are presented clearly, and are consistent with each other.

9. **Don't confuse a push for VFM with quantification**. A renewed focus on results and VFM is not a 'race to the bottom' or about doing things in the cheapest way possible. Instead, this is about improving our practice and aid effectiveness.

10. **If in doubt, get in touch with the PSST Results Hub!**

ANNEX 1: ***SUGGESTED*** INDICATORS FOR GOVERNANCE PROGRAMMING

The list below is not exhaustive, and will be updated as further research is undertaken on indicators at the programme level within each governance 'sub-theme'. Suggestions relate to programme outputs and outcomes.

For suggestions on goal level indicators, see the useful list annexed in the [Results in Fragile and Conflict-Affected States] guidance note, which provides suggested indicators against the four pillars of the Peacebuilding-Statebuilding framework.

Security & Justice	Outcome	**Oversight & Accountability**: Rating of progress against a joint national security policy / strategy / plan
		% political appointments made at senior positions in key departments
		# and % cases where S&J legislation is passed without civilian oversight and approval
		Size of military: Ratio military personnel to population size
		Military expenditure as % GDP
		Crime and public safety: % citizens who say they feel safe going out in their neighbourhood at night (disagg)
		# violent deaths recorded by hospitals / morgues
		Police: # violent crimes recorded by the police per 100,000 people
		% citizens satisfied with police response / complaints system
		% citizens who believe bribes are necessary to access police services
		Justice: % citizens satisfied with cost/quality of legal services provided
		# cases where women's rights are successfully adjudicated
		# new courts opened in rural and/or urban areas with concentrations of marginalised populations
		% judicial decisions upheld by higher courts
		% citizens using primary justice system in last year reporting satisfaction with process
		Corrections: (including pre-trail detention): Median length of stay in detention / # days between remand and first hearing
		# and % in custody over 1 year
		% trails disposed of within 6 months
		Ratio # prisoners : # beds
	Output	**Oversight and accountability**: Rating of joint national security policy / strategy / plan in terms of allocation of responsibility, authority and accountability across S&J sector
		# CSOs consulted on policy development
		Justice: # courts per 100,000 residents
		# cases where free legal advice has been provided
		% primary justice institutions using systems for recording actions and documenting decisions
		# and % disputes reported to state institutions that are referred to primary justice institutions
		# cases resolved using alternative dispute resolution mechanisms

Civil Service Reform	Outcome	**Transparency of inter-governmental fiscal relations**: PEFA Indicator 8 (Central or Sub-National) *Where no PEFA, then indicators might include:* % transfers from central government are determined by a transparent rule-based system (with criteria, formula)
		Rating of the timeliness of reliable information on the allocations to be transferred to sub-national government[5]
		% (by value) sub-national government expenditure consistent with central government fiscal reporting (by sector categories)
		Ratio ghost workers: total staff
		Effectiveness of recruitment / promotion systems: # and % of unfilled posts (Vacancy rate)
		Leavers in the last year as a percentage of the average total staff (Staff turnover)
		% people still in post after 12 months service
		Ratio sector budget allocation: expenditure: PEFA Indicator 2 *Where no PEFA, then indicators might include:* % variance between budgeted and actual expenditure (by sector)
	Output	**Effectiveness of payroll controls**: PEFA Indicator 18 *Where no PEFA, then indicators might include:* Rating of degree of integration and reconciliation between personnel records and payroll data[6]
		Time (# days) taken to make required changes to the personnel records and payroll
		# payroll audits undertaken to identify control weaknesses and/or ghost workers in the past 3 years
		% staff with job description
		% staff trained in the last 12 months to do their job (by job type)
		% complaints received acted on by Ombudsman
		# cases of disciplinary action for violations of codes of conduct per 1,000 staff

[5] Such as following the PEFA rating scale: A (“*before the start of their detailed budgeting processes*”); B (“*ahead of completing their budget proposals, so that significant changes to the proposals are still possible*”); C (“*before the start of the fiscal year, but too late for significant budget changes to be made*”); D (“*after budgets have been finalized, or earlier issued estimates are not reliable*”).

[6] Based on PEFA rating categories (using a desk or expert review).

Elections	Outcome	**Free & Fair elections**: % External observation reports stating conclusions
		% voter / political party satisfaction with conduct of the election (disagg)
		# and % political parties / parliamentarians who consider electoral commission to be a competent public body
		Participation: % voting age registered to vote (disagg)
		% voter turnout (disagg)
		# people who vote in elections supported by DFID
		S&J: % citizens expressing confidence in capacity of police to prevent and control electoral violence (disagg)
		% citizens who believe courts resolve electoral disputes fairly
		% electoral appeals concluded by courts
		# of internationally standardised electoral legal reforms implemented
		Equity: % seats in parliament held by women
		Ethnic or tribal profile of seats in parliament
		Gender profile of parliamentary candidates
	Output	**Free & fair elections**: Independent supervisory body in place - % electoral body staff trained to fulfil their specific role and responsibilities (disagg)
		# and % national election observers trained (disagg)
		Ratio national: international observers
		National election plan completion rate
		Participation: % target citizen awareness of electoral principles and procedures (disagg)
		% population aware how to vote (disagg)
		S&J: # incidents of political violence reported in national media (pre-election period; election day; post election period)
		Equity: Gender profile of polling officials

Parliament Support	Outcome	**Access to Information**: % lobby groups / CSOs/ media/ political parties stating they have accessed voting records / debate information / parliamentary information, including assets in the past 12 months
		# parliamentarians who declare assets
		Parliamentary effectiveness: % citizens satisfied with parliamentary performance (disagg)
		Length of passage of legislative reform: • # days from legislative submission to ratification • # new legislative reform bills ratified in past twelve months
		Representation: % seats in Parliament held by women
		Scrutiny quality: PEFA Indicator 27 (on annual budget law)
		PEFA Indicator 28 (on external audit reports)
	Output	**Access to Information**: # lobby groups / CSOs /media/ political parties who know how to access voting records / debate information / parliamentary information, incl. assets
		Parliamentary effectiveness: # parliamentarians trained (in what their role is and how to be effective) (disagg)
		# parliamentarians trained (in budget procedures, including scrutinising & monitoring)
		# parliamentary cttes technically trained (in scrutiny of budget, public funds, service delivery – including S&J sector)
Political Party Support	Outcome	% political parties with issue-based manifesto / codes of conduct / audited accounts
		# and % registered political parties who take actions on legal infringements
		% women / minority members of executive committees of political parties
		% political party membership given votes in internal decision-making
	Output	# and % political parties with budgeted annual plans
		% political parties with accounts available for membership or public scrutiny
		% citizens aware of right to join political parties
		% citizens who are able to mention distinguishing policies of two or more political parties

Tax / Revenue	Outcome	**Level of uniformity of tax collection**: PEFA Indicator 15 *Where no PEFA, then indicators might include:* % of tax arrears at start of fiscal year which was collected during the fiscal year (average of the last two fiscal years)
		Frequency of transfer of the tax revenues to the Treasury (daily, weekly, monthly, or longer)
		Frequency of complete reconciliation of tax assessments, collections, arrears and transfers to Treasury
		% citizens aware importance of taxation: PEFA Indicator 13 *Where no PEFA, then indicators might include:*[7] Rating of clarity and comprehensiveness of tax liabilities (desk/ expert review)
		Rating of taxpayer access to information on tax liabilities and administrative procedures (desk/ expert review)
		Rating of functional tax appeals mechanism (desk/ expert review)
		% increase in tax collection, by sector or local government (Tax collection rate)
		Tax in arrears as proportion of tax collected
		% business satisfaction with tax system
		% citizens stating complaints system is effective
		Approval of key legislative changes
	Output	Complaints system in place
		Development of key legislative changes (Tax legislation)
		Rating based on functional reviews of tax authority and tax offices (Organisational reform & capacity building)
		% increase in number of registered taxpayers (Taxpayer identification and registration)
		Higher proportion of actual collection from detected revenue in tax evasion cases (Tax enforcement)
		Average number of days to complete administrative appeals process (Efficiency measure of appeals system)
		Functional internal audit and inspection system - expert/ desk review (Audit and inspection)
		Rating of transparency of procedures for tax collection (Tax compliance and liabilities)

[7] For example, based on the PEFA "A-D" rating scale.

Corruption	Outcome	% reported cases investigated by anti corruption body (domestic or international)
		% investigated cases that lead to prosecution (domestic or international)
		% targeted civil service staff that feel safe reporting a corruption case
		% targeted citizens who believe government is committed to tackling corruption in public sector
		% national budget dedicated to anti corruption body
		% senior civil servants / parliamentarians / public office holders that declare assets according to regulations
	Output	# laws stating that corruption is a criminal offence
		% citizens who are aware of their right to access to agency info (disagg)
		% appointments to anti corruption body based on competency-based recruitment procedure
Human Rights	Outcome	% Universal Periodic Review (UPR) recommendations implemented
		Level of government accountability for service delivery: % targets achieved by human rights body
		Issue profile of CSO human rights advocacy (civil, political, economic, social etc)
		Health, education, security as a % state budget
		Health, education, security as a % household budget
	Output	# CSOs (per 100,000 persons) involved in promotion / protection of right to X
		% citizens who are aware government has legal obligation to provide basic education and other essential services for all (disagg)
		% citizens who are aware government has legal obligation to protect them from violence and abuse (disagg)
		# reported cases of domestic and/or sexual violence
Media	Outcome	Quality media law / regulator – % media owners satisfied with regulatory framework
		% target population who use media as primary source of information (disagg)
		% targeted citizens satisfied with media quality (disagg)
		Gender profile of journalists employed in national media
	Output	Media regulator capability rating
		% journalists who are aware of their responsibilities for neutrality
		# of newspapers, television, radio and internet channels, covering issues of inequality and discrimination among state and societal institutions
		# newspapers, television and radio channels, covering corruption, electoral misconduct, political violence stories

Empowerment & Accountability	Outcome	# people supported to have choice and control over their own development and to hold decision-makers to account
		PRSP = Pro Poor? / Participatory budgeting: % relevant CSOs stating they were consulted in PRSP / sector plan design / budget allocation (to an extent to which they are satisfied) (disagg)
	Output	% targeted CSOs that confirm they can obtain specified information from key public agencies
		# documented instances where PPAs are used in policy design
		% proposals from consultations used in national strategy documentation (PRSP / budget [national and sub-national] / sector programme)
		% targeted CSOs document adequate consultation in the PRSP / sector planning / budget process
		% sub-national expenditure covered by PETS
		% of targeted CSOs undertaking budget tracking
		% government expenditure online / in public libraries
		% of all national statistical publications available online