

Performance Measurement Q & A



by Mark Schacter

© October 2011

All rights reserved.

Performance Measurement Q & A

Published by:

Mark Schacter Consulting

Ottawa, Ontario, Canada

www.schacterconsulting.com

mark@schacterconsulting.com

About this paper

Following a recent engagement with a Canadian federal government client on the development of performance measures, I was asked to respond to a series of written questions from staff on elements of performance measurement that were at the top of their minds. The issues they raised were relevant not only to their Department but to public sector managers everywhere – in Canada and abroad – who confront performance measurement challenges.

This paper follows closely the text I prepared for the client. I have removed or rewritten a few elements of the original document that responded to issues particular to that Department's circumstances.

*Mark Schacter
Ottawa, Canada
October 2011*

Question 1: How to decide what indicators to include in, or exclude from, a measurement strategy?

Response:

Develop a logic model for your program and use it as a guide to developing performance measures. The outputs and outcomes in the logic model will guide you directly to areas where useful performance information is most likely to be found.

For every proposed performance measure, ask:

- Will this measure tell us something we really need to know about program performance? What would we do with the information if we had it?

- How would this information help us make decisions about the program?

- Is this measure telling us something about an area of performance over which our program has a significant degree of influence?

If you can't come up with satisfying answers to these questions, then you have no good reason to use the proposed performance measure.

Responses such as "it feels like it would be good to have this information" or "we should measure this 'just in case'" are not a basis for accepting a performance measure.

If you can establish that there is a real need for the type of information that a given performance measure will provide then you should also consider practical issues related to the level of effort and cost entailed by the measure. You may identify a measure that will generate relevant information, but if the information is very costly to gather (in terms of time and/or money) then you may have to conclude that the measure should not be used.

Question 2: Must indicators be measured at all levels of a logic model? Can we skip a level on the results chain?

Response:

A hard and fast rule is that you should not measure performance in relation to outcomes over which your program has little or no influence. This will steer you away from creating measures in relation to some levels of the logic model. (Of course, you may still – for program-management purposes separate from performance measurement – want to gather information about outcomes that are beyond your influence, but that is a separate issue.)

Apart from this, I don't believe that there are clear cut rules. Focus on collecting performance information that will help you make well-informed program management decisions and help you tell stakeholders a meaningful performance story, while bearing in mind the costs associated with performance measurement. If you think through your performance measurement strategy along these lines, and then decide that it makes sense not to collect performance information at certain levels of the logic model, so be it. The important thing is to be able to demonstrate the rationale behind your performance measurement strategy, i.e. to show why you have chosen to collect (or not to collect) certain information.

Of course, when I say that there are "no clear cut rules" I am referring to informal processes of performance measurement outside of the mandated Treasury Board Secretariat (TBS)¹ process, which does indeed contain rules (about numbers of measures, etc.) to be followed so that your formal Performance Measurement Framework (PMF) will be approved by TBS. But I am assuming that the question posed here was not about the TBS process.

Question 3: Many senior managers feel there is a "blizzard of indicators" in the department (including 260 indicators in the reporting framework submitted to TBS). Will not creating more logic models at the program level increase the number of indicators? More generally, when is enough, enough?

Response:

I want to respond to the second part of the question first – and the response is just a variation on the responses to Questions 1 and 2. I don't believe there is value in drawing a definitive line that tells you when "enough is enough". It's not helpful to say, for example, that a program with a budget of between \$50 and \$100 million dollars should have 'x' number of measures, or that you should have no more than 'y' number of measures for any given level of the logic model. For whatever rule you might make, there will always be cases where exceptions will be necessary. To be sure, there does indeed come a point where enough is enough – but I don't believe that a rule is going to tell you when you've reached that point. The challenge is to strike a balance between (a) having enough performance information to support evidence-based decision making and allow you to tell a credible performance story to the public; and, on the other hand (b) keeping the degree of information-gathering and information-analysis proportionate to your resources, needs for information and tolerance for risk.

The first part of the question asks if creating logic models at the program level will increase the number of indicators you have to work with. The short answer is that it may or may not – and this may or may not be a good thing. But what I really want to say is that the question

¹ The Treasury Board Secretariat is the Department in the Government of Canada that exercises government-wide oversight of public reporting on the performance of programs.

you have posed here may not be the right one.

The important question is whether developing program logic models and deriving performance measures from them will add value to your work as managers. If adding measures will help you make better program-management decisions and is not unduly burdensome, then adding measures would be a good thing. On the other hand, if adding measures doesn't contribute to better management, then it would make no sense to do so.

To return to the question as you phrased it – will the use of logic models increase the number of indicators? – the answer depends on how you decide to use logic models and their related indicators. One possibility is that programs are analyzed in logic-model fashion as a way to develop better results statements and measures for the formal PMF required by TBS. If so then you are not necessarily increasing the number of measures – you may just be replacing existing PMF measures with better ones. Another possibility is that you develop program logic models and derive measures from them for “informal” management purposes outside of the PMF. Because a decision to do this would be entirely at your discretion, I assume you do it if you felt it added value to your work, in which case (as I said above) the question of whether or not it increases the number of measures is beside the point.

Finally, I want to respond to the observation that managers feel lost in a “blizzard of indicators”. My reaction is not to be worried about using logic models to create new measures that may well be better than your current ones. Instead, I wonder why the “blizzard” has been allowed to rage on in the first place! Why does the overload of measures exist at all? No one – not even TBS – requires you to implement a complicated Program Activity Architecture (PAA)² that creates a burdensome set of measures. The performance-measurement harm caused by your PAA is self-imposed. You created the mess; you can fix it.

Question 4: Could you note the difference between a performance *indicator* and a performance *measure*?

Response:

The terms “performance indicator” and “performance measure” have distinct meanings, though they tend often to be used interchangeably. In general terms, the distinction is as follows. Suppose that to assess a program’s performance you need evidence of performance characteristics ‘x’ and ‘y’. ‘X’ and ‘y’ are therefore performance indicators because their presence (absence) indicates that performance is good (bad). But you need to go a step further: you also need to measure how much ‘x’ and how much ‘y’ are present. So a

² The Program Activity Architecture is a mandatory element of TBS’s policy on performance reporting. The PAA is a complete inventory of a Department or Agency’s programs, and is the basis against which Departments/Agencies state their intended results and report on performance against those results. It should be noted that the TBS policy allows Departments/Agencies considerable discretion in the way that “programs” are articulated for the purposes of the PAA.

performance indicator expresses what aspects of performance you are interested in while a performance measure describes how much you found.

Consider a skills development program aimed at reintegrating unemployed people into the workforce. A variety of phenomena might indicate whether or not the program was performing well. You might assess performance on the basis of the (a) amount of training delivered, and/or (b) trainees' perceptions of their training experience, and/or (c) impact of the training on employment. 'A', 'b' and 'c' are potential performance indicators. Then you have to ask "how much?". Indicator 'a' might be measured in terms of person/hours of training delivered, number of trainees completing courses, number of courses provided, etc. Indicator 'b' might be measured in terms of trainees' responses to scaled questions (e.g. questions answered on a scale from 1 to 5) about satisfaction with various aspects of the training. Indicator 'c' might be measured in terms of trainees' earnings in the 12 month period after the training, or the number of hours worked following training, or the percentage of trainees who report being employed following training. In short, for each performance indicator there will be at least one, and possibly more than one, performance measure.

Question 5: In one of your published papers you wrote about five basic performance measurement challenges. Could you point to a few ways to overcome these measurement challenges?

Response:

Multiple high-level outcomes. The key is to recognize, first, that multiple high-level outcomes exist. This puts you in a position to ensure that your measurement framework does not give a distorted picture of performance through undue emphasis on one high-level outcome at the expense of another. Consider, for example, a set of government programs aimed at limiting growth in greenhouse gas emissions. For argument's sake it is possible to identify at least two distinct high-level outcomes: (a) damage to the environment caused by greenhouse gases is reduced; and (b) the negative impact on the economy of changed patterns of energy use is minimized.

The implication is that performance measurement of programs aiming to limit greenhouse gas emissions needs to account for both outcomes simultaneously. It would be difficult to make a case that the programs performed well if they both reduced emissions and caused serious harm to the Canadian economy. Presumably "success" means good performance in both areas; the point is that judgment about performance must be exercised in a way that takes into account the two distinct outcomes.

Measurability. This challenge arises when there is a need to measure qualitative phenomena. Often, the immediate reaction is "that can't be measured". But it is indeed possible (e.g. through scaled survey questions) to convert qualitative phenomena (such as

awareness, learning, satisfaction, adoption of a new way of thinking) into quantitative data. The techniques that are used to do this inevitably lead to reliability risks, but when they are appropriately applied the risks can be managed.

There may be situations where a more nuanced analysis of performance is required than can be built from a few sets of regularly-reported numbers. In such cases, performance evaluation, especially impact evaluation, may be a more appropriate instrument for assessing performance than ongoing performance measurement. Impact evaluation is based on detailed research about program outcomes and their causes, incorporates multiple lines of evidence, and is better suited than ongoing performance measurement to supporting conclusions about the attribution of high-level outcomes to program outputs.

Time Lag. Although some public programs may yield significant outcomes over a short time-span, the nature of social and economic issues tackled by governments is such that in many cases it may be years before outcomes materialize. The challenge is to meet short-term performance-reporting requirements in a way that produces data of interest to decision-makers, even in the absence of evidence about high-level outcomes. This is where a well articulated logic model is useful. The immediate (and perhaps intermediate) outcomes in the logic model point to shorter-term changes that support assumptions about the occurrence, later, of high-level outcomes.

Consider the example of current efforts by the government to encourage Canadians to reduce the amount of salt in their diets. A program manager in this area might wait a long time before having anything to say about significant and widespread changes in salt consumption (and even supposing that such changes were eventually observed, there would be questions about whether they could be attributed in a meaningful way to the government's information and awareness campaign – see the discussion of attribution, below). But the program manager might be able to report, over the shorter term, significant changes in, say, awareness about the widespread presence of salt in processed foods, and about the health hazards of salt consumption. Raised awareness would be an immediate outcome in the logic model of the salt awareness campaign. Its presence there would help the program manager make a case that changes in awareness were a significant sign of good program performance even though there were still no data on longer-term changes in salt consumption.

Perverse incentives. The power of performance measurement lies in its capacity to influence behavior. People in organizations will align their behavior with what is being measured. The ideal situation occurs when performance measures motivate individual behavior that is good for the organization as a whole. But when performance measures are poorly chosen, or when undue emphasis is placed on certain measures, the incentives they create are perverse because they drive individual behavior that undermines achievement of the organization's intended outcomes.

If we continue with the example of the campaign to reduce salt consumption, suppose that heavy emphasis were placed on measuring the number of Canadians reached by awareness raising efforts (e.g. numbers of people who “like” a Facebook page or see a television advertisement or receive a pamphlet). Disproportionate emphasis on measuring the number of Canadians exposed to outreach products will motivate program managers to favor volume of outreach over value. In other words, under ideal circumstances it would be rational to expect the program to focus on, say, 100,000 people most at risk from heavy salt consumption over 1,000,000 people whose diets put them at much lower risk. But a performance measurement system that emphasized volume of outreach would motivate program staff to made the bad choice: focus on the 1,000,000 low-risk individuals instead of the 100,000 high-risk people.

The first step to minimizing the risk of perverse incentives created by performance measures is to recognize that the risk exists. For every proposed performance measure run a quick mental simulation in which you ask yourself how a reasonable person would respond if he believed his performance was to be assessed largely on the basis of that measure alone. If the result of your simulation is troubling (as it often will be!) you may be tempted to reject the proposed measure. This will not always be the right decision because just about every performance measure – even ones that have useful characteristics – carries a risk of creating perverse incentives. So it is more likely that instead of rejecting the measure you will need to think of a second measure to counterbalance the perverse impact of the first one. For example, a measure that emphasizes quantity can be counterbalanced by a measure that emphasizes quality. So to counterbalance the negative impact of counting numbers of people reached by the awareness campaign, you might want a quality-related measure such as the proportion of people reached who fall into the high-risk category.

Attribution. *The attribution problem arises because of the need to avoid measuring program performance in relation to outcomes over which the program has little or no influence. The first step in addressing the attribution problem, therefore, is to develop an argument about outcomes over which your program has a significant degree of influence (i.e. outcomes whose occurrence could reasonably be attributed to your program’s outputs) and those over which your program has little influence. A well articulated logic model is the critical instrument for organizing your thoughts around this question.*

For example, a logic model for the salt consumption awareness campaign would look something like:

- **Outputs:** *outreach products such as Facebook pages, Twitter pages, television/radio advertisements, web sites, pamphlets, official speaking engagements;*
- **Immediate Outcome:** *target group sees/reads/hears outreach products;*
- **Intermediate Outcomes:** *target group’s awareness is raised about dangers of heavy salt consumption; target group takes steps to reduce salt consumption; target group’s salt*

consumption declines over the longer-term;

*- **Ultimate Outcome:** lower incidence of disease attributable to heavy salt consumption.*

The logic model leads you directly to thinking about whether your intervention has significant influence only over exposure to the campaign, or whether it extends beyond that to awareness changes and behavioral changes. Rarely is there a definitively “right” or “wrong” answer to questions of attribution. But you have to develop an argument nonetheless, and be prepared to back it up. You might say, for example, that “we believe (for reasons x, y, z) that our program will have a significant degree of influence over awareness of health problems linked to salt consumption, but the question of whether or not heightened awareness will lead to changed behavior is affected by many factors outside our program that greatly dilute our influence. We therefore believe that the program’s performance should be assessed in relation to awareness raising among members of the target group, but not in relation to behavior change.”

About the Author



Mark Schacter lives in Ottawa, Canada, and works with Canadian and international organizations on matters related to governance, accountability, performance measurement, evaluation and strategy. He provides consulting services and professional development workshops.

Mark's publications on issues in governance and public management are widely read and used by practitioners in Canada and internationally. His publications, as well as other information about his work, can be found at www.schacterconsulting.com

Mark can be reached at mark@schacterconsulting.com or at 613.277.6777.
