Not a "Tool Kit"

Practitioner's Guide to Measuring the Performance of Public Programs

By Mark Schacter

Institute On Governance
Ottawa, Canada

www.iog.ca

Practitioner's Guide to Measuring
the Performance of Public Programs

Table of Contents

For further information, contact Mark Schacter at the Institute On Governance.
tel.: (1 613) 562 0092 ext. 237
e-mail: mschacter@iog.ca

Part 1 – Introduction

1.   Introduction

People looking for advice and assistance in public-sector performance measurement often say they want a "tool kit". I confess: I don't know what a public-sector performance-measurement tool kit would look like. In any case, if that is what you want then this Guide will disappoint you. Its basic premise is that there are no ready-made "tools", no pre-packaged "techniques", no simple "short cuts" that provide instant solutions to performance measurement problems.

The "tool" required for doing performance measurement is something that no performance measurement expert can hand to you. It is your own capacity to develop a clear, analytical understanding of the program whose performance you want to measure. There is no substitute for this.

With that in mind, this Guide aims to describe an approach to performance measurement in the public sector. The Guide will help you – no matter where in the government you work, and no matter what your sectoral specialization may be – to think your way through a wide range of performance-measurement challenges.

The central idea is that good performance measurement is an exercise in storytelling. A well developed performance framework allows you to tell a convincing story, backed by credible evidence, about the value added by your program to some particular segment of Canadian society. A good performance story addresses these three elements:

* This is what we want to achieve in Canadian society through our program. This is the area where we want to "make a difference."

* These are the steps by which we expect our program will achieve its objectives.

* This is how we know that our program is on track to achieve its objectives.

The details of the story are up to you. But if it is a good story – well written, well reasoned, and backed by credible evidence and plausible assumptions – then it will allow you to make a strong case that your program is both worthwhile (because it is pursuing outcomes that matter to Canadians) and well managed (because its progress toward achieving the intended outcomes is being carefully monitored).

The foundation of a good performance story is a detailed understanding of the program whose performance is to be measured. The first and most important step in developing a performance measurement framework is to take the program apart: analyze it, dissect it, break it up – conceptually – into its component parts.

This is not always how the task is actually approached. The first impulse of many people who are given the job of developing a performance framework is immediately to develop a set of performance indicators. This explains the desire for "tool kits". Development of performance indicators, taken out of context, sounds like a narrowly defined technical task for which there ought to be "tools."

But, in fact, the development of performance indicators ought to come at the *end* of the process of building a performance measurement framework. When performance measurement is done properly, the hardest work comes at the initial analytical stage that *prepares* you to craft a set of performance indicators.

The preliminary work, as suggested above, is all about understanding the ins and outs of the program for which you want to develop a performance measurement framework. Before you can begin thinking about indicators, you need to have a clear picture of:

- what your program aims to achieve (why are you "in business"?); and

- the steps by which you are assuming that your program will achieve its ultimate objectives (how do you propose to get from where you are today to where you want to be in the future?).

If you can develop clear, detailed and credible answers to those questions, then you have done more than half of the work required to develop a set of performance measures.

Your understanding of the program's goals and how it aims to achieve them is captured in a "logic model". The logic model provides a basis for developing a set of performance indicators. Later in this Guide we will look in detail at the logic model, and at how to make sense of the range of possible performance measures that a logic model implies.

What we can conclude, for the moment, is that performance measurement in the public sector is both breathtakingly simple and devilishly difficult. It is simple because there are really only three steps to the development of a performance measurement framework for a public program: (i) agree on the ultimate objective(s) to which the program is supposed to be contributing; (ii) create a logic model and (iii) derive performance indicators from the logic model.

It is difficult because, as we will discuss, for any given program there may be plenty of room for disagreement about ultimate objectives, and therefore about the foundation of the logic model. It is often in the nature of public programs to be simultaneously pursuing two or more high-level objectives, which sometimes may be in conflict with each other. The second difficulty comes when, having agreed on a logic model, and having used the logic model to derive a set of *possible* performance indicators, the attempt is made to choose the final set of indicators that will help you meet the challenge of monitoring the performance of the program.

This Guide devotes most of its attention to these two questions: (i) building the logic model, and (ii) dealing with the common challenges that arise as you attempt to choose a good set of performance measures.

2.   Summary

This Guide covers the three steps that are involved in the development of a performance framework for a public program:

- *first*, **agree on the ultimate objective(s) to which the program is supposed to be contributing;**

- *second*, **create a logic model that links program inputs, activities and outputs to ultimate objectives;** and

- *third*, **derive performance indicators from the logic model.**

Here is an annotated summary of the key points addressed in the Guide.

Measuring the performance of public programs is a technical exercise, but the reason for doing performance measurement is profoundly political. Performance measurement is what makes it possible for there to be strong and meaningful accountability between government Departments and the citizens they are supposed to serve. Performance measurement makes it possible for government Departments to demonstrate that they are contributing, through their programs, to outcomes that matter to Canadians *(Part 2, sections 1 and 2)*.

Performance measurement in the public sector poses special challenges. Contrary to what is normally the case in the private sector, in the public sector there is often room for ambiguity about the "bottom line". In the public sector it is not always obvious what element of program performance we ought to measure, and why we ought to measure it *(Part 2, section 3)*.

The "logic model" is a simple but powerful tool that is fundamental to the practice of performance measurement. The logic model has to be founded on a clear understanding of the ultimate outcomes to which a program is supposed to be contributing. Only once there is agreement about ultimate outcomes and about the structure of the logic model, can you then proceed to the development of performance indicators *(Part 2, sections 5 and 6)*.

Each potential performance indicator that you might derive from the logic model is a "mixed bag." Each one combines positive and negative features *(Part 3, section 2)*. There are, as a consequence, a common series of challenges that are built into the process of selecting performance indicators. These challenges are present no matter what type of public program you may be dealing with. In *Part 3, sections 2, 3, 4, 5, 6, 7 and 8,* the

Guide addresses four of the most common and important challenges you are likely to encounter in developing a performance framework:

- perverse incentives *(Part 3, sections 2 and 7)*;

- the tradeoff between meaningful results and results that are controllable (the "attribution problem"; *Part 3, sections 3 and 8*);

- the tradeoff between meaningful results and results that show meaningful change over the short term *(Part 3, section 4)*;

- the tradeoff between meaningful results and results for which data can be collected at relatively low cost and effort *(Part 3, section 5)*.

In dealing with the attribution problem, which is often the most important and difficult challenge in relation to public-sector performance measurement, it is critically important to establish the boundaries of your program's accountability. A program cannot be held accountable for outcomes over which it has little or no control *(Part 3, sections 3 and 8)*.

Unnecessary complexity is one of the most important factors threatening the successful implementation of your performance framework. Frameworks prepared for public programs often reveal a tendency to include a very high number of performance indicators and overly complicated logic models. This is a fatal error! Where performance measurement frameworks are concerned, simplicity is a virtue *(Part 3, section 9)*.

Part 2 – The Basics

1.   Why Performance Measurement Matters

What is the basic motivation for performance measurement in the public sector?

The fundamental reason why performance measurement matters to us is that it makes accountability possible, and accountability goes to the heart of our system of political governance.

Think about accountability as being part of a "basic bargain" between citizens and their government in any healthy democratic society such as Canada (**Figure 1**). Citizens grant their governments a high degree of control over their lives. Citizens allow governments to take part of their income (through taxes) and to limit their freedom (through enforcement of laws and regulations). For the most part, in a relatively healthy society such as ours, citizens are not bothered by the control exercised by government. We would prefer to live in a governed world rather than an ungoverned world. On the whole, we welcome the control over us that governments exercise.

But citizens expect their governments to be accountable to them for the ways in which they exercise power. This is the other end of the bargain: accountability *to* citizens in return for power over them. Citizens don't want to give their governments complete freedom to use their powers in any way they choose. Governments must not be allowed to abuse their powers – to use them, in other words, in ways that are contrary to the public interest. Accountability is supposed to keep governments in check, by creating pressure that causes governments to exercise power in ways that support rather than undermine the public interest. One of the methods by which governments hold themselves accountable to citizens is to monitor and report on the performance of public programs.

Before descending into the details – "inputs", "outputs", "outcomes", "logic models" and all the other technical elements that make up the practice of performance measurement – it is important to remember that the "basic bargain" is at the root of it all. It is why we care about accountability and why, therefore, we care about performance measurement.

2.   The Link Between Performance Measurement, Accountability and Results-Based Management (RBM)

The "basic bargain" means that Canadians have a right to know what governments are doing with their money, and that governments have an obligation to tell them. This principle has long been accepted in Canada, but it is important to note how views have changed over the years with respect to the kind of information that governments are obliged to provide to citizens about public programs.

For a long time, the assumption was that governments had to assure Canadians that public money had been spent in ways that complied with laws and regulations. To take a trivial example, suppose that a government department was allocated $1,000 to purchase staples. The department's accountability extended to demonstrating that it indeed used the $1,000 to buy staples (as opposed to, say, consulting services) and that the staples were procured in the proper manner (i.e. in accordance with government procurement policies).

Traditionally, this was where accountability ended. This was accountability for *compliance*. Under this type of accountability, the government assures citizens that it has spent money in ways that comply with all of the relevant laws and rules.

This type of accountability was, and remains, important. But there is a desire to go beyond that, and to ask questions about what was achieved as a result of the expenditure. Did it make a difference that mattered to Canadians?

There has been a movement toward accountability for *results* (without losing sight of the continuing relevance of accountability for compliance). To appreciate the distinction, imagine the following dialogue in connection with the purchase of $1,000 of staples:

Q.   I understand that you purchased the staples in a way that complied with all of the relevant rules and laws. But apart from that, what did you purchase the staples for?

A.   To bind together 25,000 copies of a report.

Q.   What was the purpose of the report?

A.   To examine our current public procurement process, and propose ways to make it more efficient

Q.   Why were you producing a report on this subject?

A.   If we can make public procurement run more efficiently, then less time and money will be spent on procuring goods and services, and more resources will be available for the delivery of programs and services to Canadians.

The questioner keeps pushing until he gets answers about results that matter to Canadians. Since the mid-1990s, there has been a concerted effort within the Canadian federal government to push departments and agencies toward managing for and reporting on these kinds of results. The term "results-based management" (RBM) is generally used to refer to this trend in public management in Canada.

RBM requires departments and agencies to behave as if they are continuously involved with Canadians in a dialogue much like the imaginary one, above. As **Figure 2** suggests, RBM requires that departments and agencies be ready to answer questions that Canadians might have about the impact of publicly-funded programs on the lives of ordinary citizens.

It's one thing to be focused on achieving results that matter to Canadians. Knowing whether you are actually moving in the direction of achieving those results is another matter. This is where performance measurement comes into the picture.

Performance measurement is about being able to answer a simple question:
"How do you know you are making progress toward achieving your targeted results?"

Providing a credible answer to this question means giving an answer that is backed by evidence – evidence that comes from performance measures.

This description of performance management enables us to close the circle that ties performance measurement to accountability. We said already that the conventional wisdom around the government's accountability to citizens has evolved from a single-minded focus on compliance to a broader vision of accountability that puts results at the forefront. But we can only tell a believable story about results if we are able to measure and report on our performance in a credible way – hence the need for performance measurement.

There are three key elements related to the performance of government programs, and therefore, three distinct areas where the development of performance measures would be relevant. The three areas are operational performance, financial performance and compliance.

Operational performance includes four elements:

- *relevance* – this has to do with whether or not programs make sense in view of the problems they are said to be addressing (in other words, is there a logical link between the design of the program on the one hand, and the stated objectives of the program on the other hand?);

- *effectiveness* – this has to do with whether a program is achieving what it has set out to achieve; is it generating the intended outcomes?

- *efficiency* – this concerns the cost of achieving intended outcomes; a program that achieves a given level of outcomes at a lower cost than an alternative program is performing at a higher level of efficiency;

- *integrity* – this relates to the capacity of a program to continue delivering its intended outcomes over an extended period of time; key constraining factors here relate to the ongoing availability of human and financial resources.

Financial performance covers two issues: are program spending outcomes in line with budgeted spending? and, are the financial affairs of the program being managed in accordance with sound financial management principles and controls.

Compliance performance (as mentioned already) has to do with implementing programs in ways that accord with relevant laws, authorities, policies, regulations, accepted standards of conduct, etc.

This handbook focuses on **operational performance**, and within that, on relevance and effectiveness.

3.   What is Special about Performance Measurement in the Public Sector?

Performance measurement is a more difficult and controversial undertaking in the public sector than it is in the private sector. This is because performance measurement works best when there is clarity about *what* is being measured, and *why*.

This type of clarity is characteristic of the private sector because the "bottom line" that private sector managers manage for is clear and undisputable. Private companies exist to make a profit and create wealth for their owners. There are well accepted ways to measure whether a private enterprise is achieving these objectives. You look at indicators such as profits, revenue, share-price, market share, etc.

Performance measurement in the public sector is an entirely different business, because public sector organizations exist for different reasons than do private companies. The private sector is driven by profit, so that success or failure can be measured in dollars and cents. But governments are driven by the objective of improving people's lives in ways that often can't easily be measured in dollars and cents. For public servants and the public programs they manage, there is often a lot of ambiguity over what the "bottom line" is. Ambiguity about the bottom line creates a situation that one never sees in the private sector: room for disagreement over what constitutes "results" and "performance", and therefore, room for disagreement about appropriate performance measurement.

Consider two examples, one from the private sector and one from the public sector.

Take the case of a private company that is in the business of manufacturing and marketing cigarettes. What is its "bottom line"?

The answers come to mind immediately: the company's bottom line is described in terms such as profitability, sales, market-share and share price (in the case of a publicly traded company). This uncomplicated understanding of the bottom line leads easily to a set of performance indicators.

Now consider a case from the public sector. Many of us have seen on television a series of public service announcements sponsored by Health Canada (a federal government department) that are aimed at raising viewers' awareness about the dangers of cigarette

smoking. What is the "bottom line" for this public service television ad campaign?
Would it be:

- the number of television ads produced?

- the cost of producing the ads?

- the amount of television airtime secured by the campaign?

- the number of viewers that see the ads?

- changes in awareness in the general population about the dangers of smoking?

- changes in public attitudes about smoking?

- changes in the number of people who are trying to stop smoking?

- changes in the total number of smokers in the Canadian population?

- changes in the incidence of smoking-related illnesses such as lung cancer and heart
  disease?

- changes in costs incurred by the health-case system as a result of smoking related
  illnesses?

Unlike what we observed in the private sector example, in this case the answer is not
obvious. Any one of these items could conceivably be appropriate descriptions of some
aspect of the "bottom line" of the television ad campaign. Each item, therefore, is
conceivably a potential performance indicator. But it will require further analysis and
judgment to narrow down the list and decide which ones are indeed appropriate
descriptions of the bottom line and, therefore, appropriate performance measures.

In the private sector example, we did not hesitate. We accept automatically that
indicators such as profitability and sales are tightly linked to the "bottom line" and
provide an objective and reasonable basis for measuring the performance of the company.

In the case of the public service ad campaign, the answers may at first seem simple, but
the more we think about it, the more we see that the appropriate selection of performance
indicators requires much more judgment and analysis than in the private sector case.
Yes, the ultimate objective being served by the advertisements is to have fewer people
getting sick and/or dying as a result of smoking cigarettes. But for the purposes of
measuring the performance of the ad campaign, is this ultimate objective really our
"bottom line." Would it make sense to hold the managers of this ad campaign
accountable for changes in rates of lung cancer and heart disease in Canada? Would it
make sense to hold them accountable for changes in the number of Canadians who
smoke, or for changes in societal attitudes about smoking?

Where do we draw the bottom line in this case? Where does the accountability of the publicly-funded program end? At what point can the managers of the ad campaign say, with good reason, "We are not accountable for *that*. It isn't fair to judge our performance on the basis of *that* indicator."

As we said, the answers to these questions are not obvious. But we can't leave it at that. Unraveling this puzzle is the central challenge of doing performance measurement in the public sector. We need an approach to addressing this challenge. We need to be able to analyze our way through the question of where accountability ends and of what constitutes an appropriate set of performance measures for a public sector program. We need a general way of thinking about this problem that can apply broadly, across a wide range of public sector activities.

We are going to address these questions later on in this Guide. Before doing so, we need to sort out some basic questions of vocabulary.

4.   Inputs, Activities, Outputs and Outcomes

The example of the public-service advertisement campaign reminds us of the need for clarity in relation to terms such as "input", "activity" "output" and "outcome". These words are the basic language of performance measurement. Traditionally, they have been used in connection with physical processes such as manufacturing rather than with public programs or public policy. As a result, they can most easily be explained by using the metaphor of a production process at a factory.

Inputs are the raw material of the production process. If we were manufacturing a car, our inputs would include steel and glass and rubber, as well as the labor provided by people working on the assembly line.

Activities include everything that is done in order to transform the raw materials into the final product. In an automobile manufacturing plant, activities would include the assembly of the various components of the automobile.

Outputs are the finished products that role off of our assembly line. In this case, our output is an automobile.

Outcomes are what happen outside of the production process as a direct or indirect consequence of our output. From the perspective of the owner of the automobile company, a key outcome is the financial success of the company. From the perspective of automobile users, a key outcome is transportation – people are able to move from place to place with relative ease.

In the case of public programs (**Figure 3**), *inputs* are typically people and money. *Activities* comprise all of the things that people involved in the design and/or delivery of

a public program do that are related to the program. So, activities might include production of reports, preparation of analyses and research, consultation with stakeholders, visits to program sites, etc. *Outputs* are the products or services that the program ultimately makes available to a target group. *Outcomes* are what happen out in Canadian society as a result of the program outputs. Outcomes in the context of public programs are normally intended to be *changes for the better* in Canadian society. Outcomes at this level – i.e. contributions to improving the lives of Canadians – are normally the reason why public programs exist in the first place.[1]

As Figure 3 indicates, public servants "keep busy" by occupying themselves with activities and outputs. But it is only when outcomes begin to occur that public programs can make claims about "making a difference" to the lives of Canadians.

If we assume that most public servants would rather be known for "making a difference" than for "keeping busy", then it follows that performance indicators related to outcomes are more meaningful than those related to outputs, activities or inputs. But it is also the case – as we will examine in detail later in this Guide – that outcome indicators also tend to be harder to measure, and create the greatest challenges in terms of "attribution" – i.e. in terms of making a link between the thing being measured by the performance indicator and the program itself.

In other words, the things that we care about the most from the perspective of public programs are also the things that create the most difficult performance measurement challenges. Not surprisingly, therefore, when government organizations are required to implement performance measurement, they tend to gravitate away from performance measures related to outcomes and toward measures related to activities and outputs. This points us to what is perhaps the most fundamental challenge in public sector performance measurement: the gap between what is measurable and what is meaningful. We will address this challenge later in the handbook. But first, we need to examine a simple but powerful tool – the logic model – which is used to integrate inputs, activities, outputs and outcomes into a meaningful and compelling "story" about the performance of a public program.

5.   The Logic Model

The logic model is fundamental to the practice of performance measurement.

---

[1] This illustrates an important distinction between the ultimate *intended* purpose of public as opposed to private activities. The activities of private enterprises may generate social benefits, but these are either unintended or at best a secondary intention of the enterprise owner. The primary interest of the private entrepreneur is to generate private benefits by earning a return on investment. By contrast, the main purpose of public activities is to generate social benefits. Social benefits, by their nature, are likely to be diffuse, contain an important qualitative element, and be the result not only of a particular public program but also a variety of other unrelated factors. This explains why measuring performance in relation to ultimate outcomes is much more challenging in the public sector than in the private sector.

As a manager of a public program, part of your accountability obligation is to be able to tell your stakeholders a believable story about why your program is put together in the way that it is. Why is 'A' expected to lead to 'B', 'B' to 'C', 'C' to 'D', and so forth? In the case of the anti-smoking television ads, it might be reasonable for stakeholders to ask questions about the assumed link between producing advertisements and lower rates of lung cancer. How do you get from the one to the other?

The logic model is your answer to these questions. It is a vision of "how the world works" from the perspective of your particular program. It is the backbone of the convincing story you need to tell about your program, if you want to be able to measure its performance.

A logic model ties together, in a logical chain, the inputs, activities, outputs and outcomes relevant to the program for which you are developing a performance framework. **Figure 4** provides a generic illustration of a logic model. **Figure 5** illustrates what the logic model might look like for the anti-smoking advertisement campaign on television. A narrative explanation of the logic model might go something like this:

> We are going to produce a series of television advertisements (*activity)* that convey messages about the harmfulness of cigarette smoking. We will adopt an ad-buying strategy (*activity*) such that when the ads appear on television (*output*) we are likely to reach our target audience (*immediate outcome*). As well, we will carefully design and pre-test the ads (*activities*) to increase the probability that attitudes of the people who see the ads will be affected in ways that make smoking appear less attractive (*intermediate outcome*). As a result of the impact that the ads will have on people's attitudes, we assume that the behavior of the people who see the ads will be affected and that they will reduce their consumption of cigarettes, or quit smoking, or not start smoking (*intermediate outcome*). Because less people are smoking than would otherwise have been the case, we will eventually see a reduction in smoking-related diseases such as heart disease and lung cancer (*ultimate outcome*).

A logic model forces you to think through, in a systematic way, what your program is trying to accomplish and the steps by which you believe your program will achieve its objectives. Because measurement of program performance is tied to program objectives, and because the logic model articulates the series of steps by which a program is intended to achieve its objectives, the logic model is also the foundation for developing a set of performance measures. *If a logic model is well done, a set of appropriate performance measures emerges from it almost automatically* – a point that we will address in more detail below.

It is also important to note – although we will not address these subjects in this handbook – that the logic model is also linked closely to *risk management* and *program evaluation*. This is so because a logic model reveals the assumptions on which a program is based. In the case of the anti-smoking advertisements, the logic model helps us see that the following are some of the program's key assumptions:

- 1 – the ads will be seen by a minimum acceptable number of people;

- 2 – the ads will have the intended impact on people's attitudes about smoking;

- 3 – changes in attitudes about smoking will lead to less smoking;

- 4 – reduced smoking will lead to reduced rates of lung and heart disease.

The propositions contained in assumptions 1, 2 and 3 are far from certain. There is a significant possibility that too few people will see the ads, and/or that the ads will not have the intended impact on people, and/or that they will not persuade many people to reduce, quit or not begin smoking. (By contrast, the proposition in assumption 4 is robust. A substantial body of scientific evidence allows us to make a firm connection between smoking and heart and lung disease.) The fact that some of the key assumptions may turn out to be incorrect is what makes this program *risky*. If any one of the assumptions fails, then the program will fail to achieve its objectives. Does this mean the program should be scrapped? Probably not. Risk is an inherent feature of just about any human undertaking. But we do need to understand the risks we are taking, so that we can prepare for them. This is where the logic model helps. By laying bare the assumptions built into our program, it provides a basis for *identifying* key risk factors, and for developing plans aimed at managing and minimizing the risks. (Most risks can never be entirely eliminated. Virtually all public programs are inherently risky.)

Similarly, by laying bare the assumptions underlying the program, the logic model provides a basis for program evaluation, an activity that normally takes place at the end of a project ("summative evaluation") or after a program has been underway for a number of years ("formative evaluation"). An evaluation looks in depth at a program's impact (or lack thereof) and seeks to understand why a program has succeeded or failed. Evaluation studies will often probe the original assumptions, evident from the logic model, that underlay the program's design and implementation.

6.   **Be Clear About Ultimate Outcomes**

The most important thing that a logic model does is force you to think about the ultimate outcomes to which your program is supposed to be contributing. A logic model must be built from the "bottom up." Be clear, first, about the outcomes that are at the end of the logic model. Then work backwards through outputs, activities and inputs. The visual representation of the links between the various elements of the model is a reminder that inputs, activities and outputs only make sense in relation to the outcomes that they are supposed to be influencing. The model compels you to address the question, "What is our program aiming to contribute to out in Canadian society that is going to make a positive difference to people's lives?"

The logic model leads us, therefore, to the most fundamental principle of performance measurement: you cannot do a good job of performance measurement in the absence of

agreement on high-level outcomes. High-level outcomes drive the design of your logic model, which in turn drives the selection of your performance indicators (**Figure 6**).

Simply put, if there isn't agreement around the high-level outcomes to which a program is supposed to be contributing, then there can't possibly be agreement on how to measure the performance of the program. A story from British politics provides a vivid example:

> A professor of British politics had written his doctoral thesis in the early 1960s on the British Housing Act of 1957. About ten years ago, he decided to refresh his research on the topic, and went back to interview the man who had been the minister responsible, Mr. Enoch Powell. The professor begin his interview by noting that everyone agreed that the Housing Act in question had been a failure. He was about to follow up what he believed to be an uncontroversial statement of fact, when Powell cut him short. "Whatever do you mean that the Act was a failure?" he asked. Startled, the professor replied that the Act's stated objective was to build 300,000 houses a year, and that in no year when the Act was in force had anything like that number of houses actually been built. "My dear chap," Powell replied, "the objective of the Housing Act of 1957 was to get rid of housing as a topic of political controversy. It was so successful that housing did not surface as a contentious political question for the two or three subsequent General Elections. The Housing Act was an unqualified success."[2]

Radically different assumptions about outcomes prevented Powell and the researcher from having a meaningful conversation about performance. For Powell, the major objectives of the Act were political. The key performance measure, for him, was the degree to which public views about housing helped or hurt his party's chances at election time. For the researcher – whose perspective was more attuned to that of a public servant – the major objectives were societal, related to increasing the stock of low cost housing. For him, the key performance measure had to do with the number of houses built.

The point here is not to make judgments about whose perspective is "right". We can appreciate how each point of view might make sense. It all depends on your assumptions. The lesson is that a performance-measurement framework for any public program is only going to be meaningful *in relation to assumptions about ultimate outcomes*. For any given program, different assumptions about ultimate outcomes imply different logic models and, therefore, different sets of performance indicators.

Reaching agreement on the high-level outcomes to which a program is supposed to contribute is not only the most fundamental step to developing a performance measurement framework, but is often also the most difficult step.

---

[2] This anecdote is extracted from "Results are in the Eye of the Beholder," by Brian Lee Crowley, *The Alternative Network*, Vol. 2.4.

Part 3 – Key Challenges to Developing a Performance Framework

No matter what type of program you are involved in, no matter which government department you work for, the challenges that you will face in developing and working with performance measures will be remarkably similar.

In this section of the Guide, we outline four major challenges that are common "across the board" to the design and implementation of performance measurement in the public sector. We will also sketch out some approaches to dealing with these challenges.

1. Every Performance Indicator is a "Mixed Bag"

We observed above that once we have developed a logic model, it becomes a relatively simple matter to generate a list of possible performance indicators for your program. **Table 1** shows how this would work for the logic model we developed for the anti-smoking television advertising campaign.

Table 1

Layer of Logic Model	Related Performance Indicator	Positive Features of Indicator	Negative Features of Indicator
Input	actual program spending in relation to budget	obtaining data is easy and inexpensive	weak relationship to outcomes
Activity	number of ads produced	obtaining data is easy and inexpensive	weak relationship to outcomes
Output	ads appear on television	obtaining data is easy and inexpensive	weak relationship to ultimate outcomes
Outcome	number of people who see ads	relatively easy to obtain data	weak relationship to ultimate outcomes
Outcome	influence of ads on viewers' attitudes	moderate relationship with ultimate outcomes	effort/cost of obtaining data
Outcome	changes in level of smoking	strong relationship with ultimate outcomes	effort/cost of obtaining data; uncertain cause & effect
Outcome	changes in smoking-related diseases	reflects the ultimate outcome	uncertain cause & effect

Notice two important things in Table 1. First, it is indeed easy to generate a list of possible performance indicators from the logic model. Each layer of the logic model – each input, activity, output and outcome – suggests a performance indicator. Second, each performance indicator is a "mixed bag." Each has positive and negative features.

Take for example an indicator related to the "input" layer of the logic model. Financial resources are an input; the related performance indicator has to do with whether actual

program spending is in line with budgeted spending. This is a "good" indicator because it is easy to manage. The difficulty and cost of obtaining the related data are low. It is also good because it generates information that is required for day-to-day management. But it is a "bad" indicator because it bears little relationship to the ultimate purpose of the advertising campaign. Measuring actual spending performance in relation to the budget tells you nothing about whether the advertising campaign is contributing to the ultimate outcome: reduced incidence of smoking-related diseases.

Now take an indicator derived from the outcome layer of the logic model. Reducing the incidence of smoking-related health problems is the ultimate outcome identified in the model. The related performance measure is the actual incidence of smoking related diseases. This is a "good" indicator because it is identical to the ultimate outcome to which our program aims to contribute. It is a "bad" indicator because it measures something over which the program has little (if any) direct control. The program is, we assume, making a contribution to reducing levels of smoking-related diseases. But would it be fair to hold the management of the program accountable for changes in rates of heart and lung disease in Canada? Certainly not! In order for us to use data about heart and lung disease as a basis for judgments about the performance of the program, we would need to see a clear, direct cause-and-effect relationship between the advertisements and rates of heart and lung disease. If we cannot establish such a relationship – and in this case we clearly can't – then it doesn't make sense to use data on heart and lung disease as performance measures for the program. (Much of the rest of this handbook is devoted to considering questions of cause-and-effect, and their relationship to performance measurement.)

As you move down the logic model from inputs to ultimate outcomes, you observe a general pattern in the mix of "good" and "bad" attributes of performance indicators. Near the top of the logic model, "good" features generally relate to the ease and low cost of data collection (data on inputs, activities and outputs is often available in-house) and to the relevance of the indicators to internal, short-term management matters, while "bad" features generally relate to the remoteness of the indicators from the ultimate outcomes of the program. (Indicators at the input, activity and output level generally tell you little about whether the program is helping "make a difference" to Canadians).

Near the bottom of the logic model, "good" features generally relate to the relevance of the performance indicators to ultimate outcomes. (Indicators at the outcome level give you a good sense of whether the program is making a difference to Canadians.) "Bad" features often have to do with the ease and cost of data collection (data on outcomes is often not available in-house) and with uncertainty about cause-and-effect.

Every performance indicator, as we said, is a "mixed bag". Each one has a combination of positive and negative attributes. This observation leads us naturally to a discussion of the challenges that are built in to the process of selecting performance indicators. In the following sections of the Guide, we address four of the most common and important challenges you are will encounter in developing a performance framework:

- perverse incentives;

- the tradeoff between meaningful results and results that are controllable (the "attribution problem");

- the tradeoff between meaningful results and results that show meaningful change over the short term;

- the tradeoff between meaningful results and results for which data can be collected at relatively low cost and effort.

2.   Challenge No. 1 – "Perverse Incentives"

Imagine a government-run service program – say, for example, the issuance of drivers licenses. Suppose that the program's top management had decided that one of the program's high-level outcomes (i.e. at outcome close to the bottom end of the logic model) must be "client satisfaction". The mission statement of the "Drivers License Office" declares therefore that clients will be treated courteously, and that licenses will be processed as quickly as possible and with a minimum of errors.

Now suppose that management developed a performance measurement framework that emphasized only the speed with which clients were processed. Among other things, the performance evaluations of the staff were tied tightly to the number of license applications that were processed every day.

What is the likely result of this approach to performance measurement? Staff, conscious that management is closely monitoring them for speed, are likely to forget about being courteous to clients (being courteous takes time) and are likely to make more than an acceptable number of mistakes in processing licenses (being accurate takes time).

Management has created a "perverse incentive". They have set up a performance framework that is motivating staff to behave in ways that are opposed to some of the Office's formal, high-level outcomes.

All federal public servants are familiar with the rush to disburse funds that occurs during the period from January through March. This is a powerful example of a perverse incentive. It exists because "lapsing" the smallest possible amount of budgeted program resources at the end of the fiscal year has come to be regarded as a key indicator of "good management". The (often incorrect) assumption is that the smaller the lapsed amount, the better the manager. Managers, even if they have good reasons for not using their entire allocation, worry about the harm to their reputation if they leave funds unspent. But of course, spending money by March 31 has no necessary connection with the high-level outcomes that programs are serving. It has nothing to do with "making a difference" to Canadians. End-of-year spending pressure pushes program managers to

act in ways that have nothing to do with results that Canadians care about. It is a classic example of a perverse incentive. (See **Box 1** for another example.)

The lesson we learn from perverse incentives is that there can be nasty consequences for organizational performance if you choose a faulty set of performance measures. Performance measures send a powerful signal in an organization. They say, "We are paying attention to this!" When career advancement and personal reputation are tied meeting performance targets, people will respond as you might expect them to. If you measure the wrong things, people will often respond by doing the wrong things.

Box 1 – Perverse Incentive: Hospital Report Cards

Critics say that hospital "report cards" encourage hospitals to reject sick patients, who are harder to treat and drag down performance scores.

Recent research suggests that the critics have a point. Researchers looked at the experience of elderly heart-attack victims in New York and Pennsylvania, states which publish mortality rates for coronary bypass surgery for particular hospitals and surgeons. They found that the report cards encouraged hospitals to do bypass surgery on relatively healthy patients. There is also evidence that hospitals were less likely to accept sicker patients for the surgery.

The researchers concluded that the report cards contributed to increased costs for the publicly-funded Medicare system. Hospitals did bypasses on relatively healthy patients who could equally have received cheaper angioplasty surgery. On the other hand, sicker patients who didn't receive bypasses wound up back in the hospital with more heart attacks.

The researchers observed that the report cards were not necessarily harmful to health, but that scoring methods should be changed in order to reduce hospitals' incentives to produce perverse outcomes.

Source: Business Week, April 1, 2002

---

3. Challenge No. 2 – The Tradeoff between Meaningful Results and Results that are Controllable (or the "Attribution Problem")

The further you move down the logic model toward results out in society – i.e. toward the kinds of results that Canadians care about most – the further you move away from results that are within the control of your program. The closer you get to the bottom of the logic model, the more difficult it becomes to make believable claims about responsibility for outcomes. This is often referred to as the "attribution problem".

In the case of our anti-smoking advertisements on television, the ultimate outcome to which the government is trying to contribute is reduced rates of smoking-related diseases. This is the outcome that is most important to Canadians.

But what if we observed, after the advertisements had been running for a year, that rates of smoking-related diseases were declining. Would it make sense to attribute this positive development to our advertisements? Should we take the credit? Conversely, what if there was no significant decline in disease after a year's time? Worse, what if the incidence of smoking-related diseases increased over the period. Would it make sense to conclude that our advertisements were ineffective (or counterproductive)? Should we take the blame?

Common sense tells us that the answer to these questions is "No!" Many powerful factors apart from our television advertisements affect the presence of heart disease, lung cancer and other smoking-related ailments in the Canadian population. Our efforts are contributing (we assume) to improving the situation. (And our logic model allows us to explain exactly how we think our advertisements are contributing to making a difference.) But so many other important factors are at play that it is practically impossible to make a direct causal attribution between our advertisements and the number of people suffering from smoking-related health problems.

For all practical purposes, our program has no control over the ultimate desired outcome: fewer people suffering from smoking related diseases. It makes no sense to judge the performance of a program on the basis of something over which it has no control. So "rates of smoking-related disease" is not an appropriate performance measure for our anti-smoking advertisement campaign.

What if we moved one step up the logic model, and looked at levels of smoking. Would it make sense to attribute decreases/increases in the numbers of Canadians who are smoking to the success/failure of our advertising campaign. Or are there too many other factors at play? Do we have enough control over this particular outcome to justify using it as a basis for measuring our performance? Probably not, although the point is of course a debatable one.
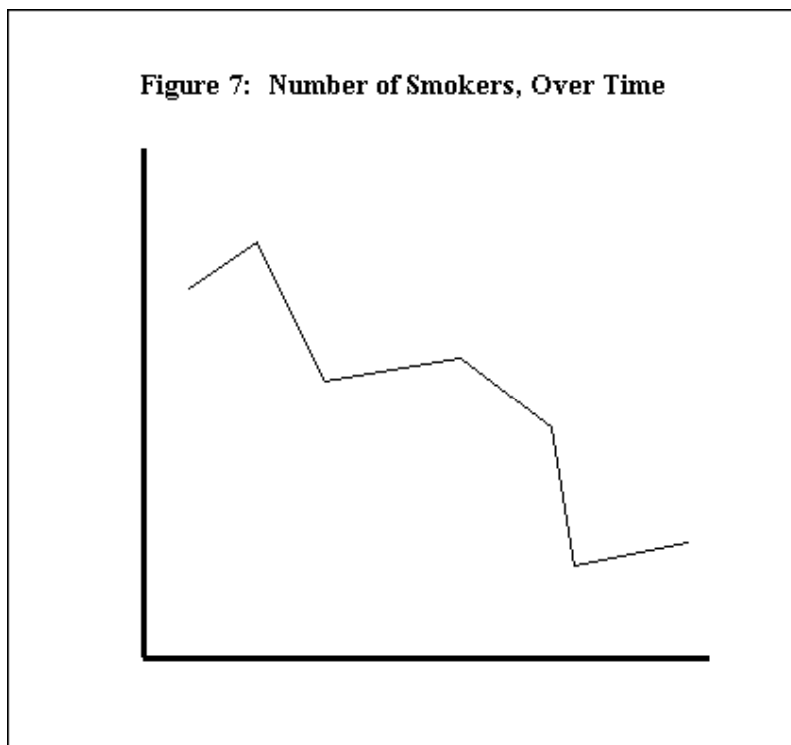
Where on the logic model do we draw the line between what we can and cannot be held accountable for? We will address this question later in the Guide. What we can conclude now is that setting the boundaries of accountability within the logic model will often be a question of informed judgment, rather than fact.

What we *can* say with certainty is that there will almost always be a tradeoff to address. We care most about the outcomes that are often least suitable to be used as performance measures; conversely, the outcomes and outputs that are most suitable to be used as performance measures (because they relate to phenomena over which our programs have significant control) tend to be further removed from the results we care about the most.

4.   Challenge No. 3 – The Tradeoff between Meaningful Results and Results that Show Meaningful Change over the Short Term

Performance reporting in the public sector proceeds in a neat, annual rhythm. The outcomes that you must report on almost never do. Let's assume, for example, that we agree to use changes in the number of smokers in Canada as a performance indicator for our anti-smoking advertising campaign.

On the one hand, this is good. We have a performance indicator that is closely related to our ultimate desired outcome. But on the other hand (remember, we said that every performance measure is a mixed bag of "good" and "bad"), we have a problem. You have to produce annual performance reports which will show – you hope – that your program is making regular progress. Unfortunately, in "number of smokers" you have chosen a performance indicator that may not help you make your case, because meaningful changes may not be evident on a year-to-year basis. Imagine a graph plotting the number of smokers in Canada over an extended period of time. Suppose that the graph looked like **Figure 7**. The data here are "noisy" over the short-term. Short-term patterns may be spurious and misleading. In our example, the overall long-term trend is downward – a good thing. But if you only looked at certain annual periods in isolation, you might wrongly conclude that the number of smokers was trending upward. It is only over the longer term (perhaps periods of five years or more) that we might be able to make meaningful statements about trends, upward or downward, in numbers of smokers.

Figure 7: Number of Smokers, Over Time

This is typical of results that are close to the end of the logic chain; they often reflect major societal changes that may take a long time to materialize. So this kind of indicator is not well suited to the task of demonstrating to stakeholders, on an annual basis, that you are making progress.

5.   Challenge No. 4 – The Tradeoff between Meaningful Results and Results for which Data can be Collected at Relatively Low Cost and Effort.

If we look at the group of possible indicators implied by the logic model for the anti-smoking campaign, we see a wide range in terms of the ease and cost of the related data-gathering.

Consider some of the indicators near the top end of the logical model – indicators related to inputs, activities and outputs. For the most part, the data are available cheaply (they are in-house) and require relatively little effort to collect.

The picture changes as we move down the logic model to outcomes. The first thing we notice is that data related to outcomes – numbers of people who see the ads, impact of the ads on viewers' attitudes, changes in numbers of smokers, changes in incidence of smoking-related diseases – will generally not be available in-house. This suggests more time and effort to collect the data than would be the case for data related to inputs, activities and outputs. There may also be significant financial implications related to the collection of outcome data. For example, in order to assess the impact of the ads on viewers' attitudes about smoking, it might be necessary to commission special focus groups, surveys or evaluation studies.

The general pattern is that as you move down the logic model toward measuring outcomes, you face increased costs and effort related to gathering the data that are tied to the performance indicators.

6.   Dealing with the Challenges – A Question of Balance

Later in the Guide we will look in detail at how to address some of these challenges. But first, consider this general point. Dealing successfully with the challenges requires you to adopt an overall "mindset" in relation to developing a performance framework (which is a logic model and its related group of performance indicators). You need to accept from the outset that not only – as we have already observed – will no single performance indicator be perfect, but also that no performance framework will ever be perfect. The Auditor General of British Columbia, in a 1997 report, captured the point nicely:

> Achieving the best and most productive mix of accountability and performance management is a question of balance. No complex system of this nature will satisfy all its participants all the time.

No one performance indicator will tell the story about your program; nor will your performance framework tell a perfectly complete story about your program. But a well developed framework – a well crafted logic model and related set of performance indicators – will allow you to tell a reasonably convincing and credible story. The

inevitable weaknesses in some parts of your framework will be balanced out by the strengths of others. The whole will be more than the sum of its parts.

So there is a general approach to addressing the challenges and tradeoffs that are an unavoidable part of measuring the performance of public programs. The key is to recognize that:

- perfection is not possible;

- there are tradeoffs to be made;

- it's up to *you* to decide how to make those tradeoffs;

- you need to be aware, as you choose performance indicators, that you are making choices; you are trading off certain benefits against others;

- only when you are clear in your own mind about the tradeoffs you have made, and why you have made them, can you explain to others how the pieces of your performance framework fit together, and why the whole thing makes sense.

The point about tradeoffs is illustrated in **Figures 8, 9 and 10,** which present an analysis of the performance indicators implied by the logic model for the anti smoking campaign. Each indicator is scored on a scale of "high", "medium" and "low" in relation to five qualities:

- meaningfulness of the indicator in relation to ultimate outcomes ("high" is ideal)

- difficulty of data collection ("low" is ideal);

- level of program control over the measured phenomenon ("high" is ideal)

- risk of creating a perverse incentive ("low" is ideal)

- meaningful change on an annual basis ("high" is ideal)

Although you may not agree exactly with the placement of the arrows in the illustrations (the scoring of each indicator is a matter of judgment, not fact), the illustrations make the point that every indicator includes tradeoffs among positive and negative attributes.

Having briefly described a general approach to addressing the challenges and tradeoffs inherent in building a performance framework, let's now take a more detailed look at two the challenges: perverse incentives and the attribution problem.

7.   Dealing with Perverse Incentives

The most effective first step in dealing with the risk of perverse incentives is simply to be alert to the fact that the risk exists. The risk will always be present. If you ignore it, it is certain to harm your program in one way or another. If you are aware of it, you can take measures to minimize or neutralize its impact.

Be mindful of the potential of every performance indicator to influence, for better or for worse, the behavior of staff and management. When you are at the preliminary stage of choosing a set of performance indicators, try to think about the perverse risks that may be built in to each one, and then make a decision about whether the risk is worth taking.

Consider again the example of the "Drivers License Office". Management chose a performance indicator – the rate at which clients are processed – that carried a strong risk of generating a perverse incentive. On the other hand, there were good reasons for choosing this indicator. It provided management with important information on the efficiency of staff, and was relevant to a significant element of the Office's overall mission of client satisfaction.

The indicator carried a strong risk of creating a perverse incentive, but the risk was worth taking because the indicator also had important positive attributes. But some advance analysis of possible perverse effects of this indicator would have allowed management to take precautions. The error that management made, in our example, was an error of emphasis. Management attached too much weight to one indicator alone, and the result was a foregone conclusion. In organizations, people behave in ways that conform to the things being measured. In our example, they focused on speed, at the expense of courtesy and accuracy.

What if management had taken a more balanced approach? What if its performance framework had given equal weighting to indicators of speed, courtesy and accuracy? It's unlikely, under those circumstances, that we would have seen such a single-minded focus by staff on speed.

This points to a simple and important strategy for minimizing the risk of perverse incentives. Once the risk has been identified, and if it is decided that the risk is worth taking, then look for other indicators flowing from your logic model that might counter-balance the perverse impact of the risky indicator.

As well, there is no substitute for constantly monitoring the performance framework for signs that it is having a perverse impact on the behavior of staff or management. If there are signs of perverse effects, then management must be prepared to adjust the framework (e.g. by eliminating the indicator in question, or by adding others to counter-balance its perverse influence).

8.   Dealing with the Attribution Problem

This is the most difficult challenge you will have to deal with as you develop your performance measurement framework. What level of outcomes is it reasonable to attribute to your program? At what point on the logic model does your accountability end? How do we explain and justify to stakeholders our decisions about where to "draw the line." Dealing effectively with the attribution problem requires thoughtful answers to these questions.

The attribution problem is difficult not only from a technical perspective but also from a "public relations" or political perspective. Many of the key stakeholders who will review your performance information – your Minister, parliamentarians, news media – will be expecting you to hold yourself accountable for ultimate outcomes.

As suggested, accountability at the program level for high level outcomes will in most cases be an unreasonable notion. It would be nonsense to hold the manager of the anti-smoking television ad campaign accountable for rates of smoking induced diseases in Canada. Too many other factors (social, demographic, economic, environmental) are affecting the same outcome.

The error of holding program managers accountable for high-level outcomes is an important point that risks getting lost as the federal government puts increasing emphasis on "managing for results". No reasonable person would argue with the importance of managing for results. But there is a distinction to be made between "managing for results" and being required to *attribute results* – high level, ultimate results – to your programs[3]. The point cannot be repeated too many times: programs should not be held accountable for, and their performance should not be judged on the basis of, outcomes over which they have little or no control. As the Office of the Auditor General has observed,

> The manager might have indeed done all that could be expected, but the results were not achieved due to circumstances beyond his or her influence. To encourage and support managing for results, we need a new view of accountability that acknowledges this more complex management world. Attribution here is a real problem.[4]

In view of the attribution problem, how do you justify your program from a results perspective? You tell a good performance story. In other words, you must:

---

[3] It is interesting to note the exceptional cases where it *does* make sense to hold program managers accountable for high-level outcomes. In immunization programs, for example, you have a direct link between the program output (injections delivered) and the ultimate outcome (reduced incidence of the disease in question). In an immunization program the link between output and ultimate outcome is clear, direct and non-controversial. Very few public programs fit this model.

[4] "Addressing Attribution Through Contribution Analysis: using Performance Measures Sensibly," by John Mayne, Ottawa: Office of the Auditor General, 1999.

- make a convincing argument (founded in your logic model and its related set of performance indicators) which demonstrates that your program is *likely to contribute* to ultimate outcomes (or results);

- make a convincing case that your program is being *managed for results*, even if you cannot always *prove* that there is a direct causal link between your program's outputs and the desired final results;

- find a way to demonstrate that your program is achieving results at some meaningful level (intermediate outcomes), even if you can't show a direct link between your program and the ultimate outcomes.

(i) Distinguish between Two Types of Results Indicators

In dealing with the attribution problem, your performance story needs to make a clear distinction between managing for results and taking accountability for results. In order to do that you have to make a distinction (one that is often ignored) between two kinds of indicators.

On the one hand, there are indicators measuring things that are a *fair reflection of program performance*. These are suitable to be used as performance indicators for your program. On the other hand, there are indicators measuring things that are *related to program performance*. As we will elaborate below, these should be a part of your performance story, but they are not suitable to be used as performance indicators.

Indicators that are a fair reflection of program performance measure things (typically inputs, activities, outputs, immediate outcomes and intermediate outcomes) over which your program (i) has some reasonable degree of control and which (ii) have a logical connection to the ultimate results.

These indicators enable you to say, believably, "if our program is successful at this intermediate level, then it's reasonable to conclude that it is contributing to achieving the ultimate outcomes identified in the logic model." These indicators reflect the limits of your "comfort zone" with respect to attribution. You would feel that it was reasonable to attribute to your program the results (or lack of results) that were linked to these indicators.

Indicators that are related to program performance measure things (typically intermediate and ultimate outcomes) over which your program has no, or virtually no, control. They are nevertheless things that you must monitor, because they have profound relevance to the design and the implementation of your program. *These are the results that you are managing for even if you can't control them.* These indicators are outside the limits of your attribution "comfort zone." You would not feel that it was reasonable to

attribute to your program the results (or absence thereof) that were linked to these indicators.

Figure 11 illustrates how this would work in the case of our anti-smoking television campaign. *Indicators that are a fair reflection of program performance* are shown above the line. They are suitable to be used as performance indicators because they provide information about results:

- over which the program has some reasonable level of control;

- which have a logical link to the desired ultimate outcomes;

- which show meaningful changes on an annual basis;

- and which provide credible evidence that you are contributing to ultimate outcomes.

Indicators that are related to program performance are shown below the line. They address the issues that Canadians care about most in relation to your program: number of smokers, and the incidence of smoking-related diseases. But the attribution problem – as well as the problem that these results will generally not show meaningful movement within the annual reporting timeframe – precludes their use as performance indicators for the program.

(ii) Both Types of Indicators Belong in Your Reporting Package

At the end of the day, you want to be able to tell a story about how the indicators "above the line" are affecting the indicators "below the line." That's why even though only the first type of indicator is suitable for use as a program performance indicator, *both* types of indicators belong in your reporting package. Information generated "below the line" must be part of your reporting because:

- you and your stakeholders need to know whether progress is being made toward the ultimate outcomes;

- the ultimate outcomes shape the context within which the program operates; they are an important basis for decisions about program design and implementation;

- information from indicators "below the line" provides a basis for making judgments about the logic of the program (e.g. given the level and trends that we see in the data related to ultimate outcomes, does the design and the implementation of the program make sense?).

Think about this in the context of the anti-smoking television ads. In Figure 11, we have placed "number of smokers" and "incidence of smoking-related diseases" below the line.

We are not holding ourselves accountable for these outcomes, and are not judging the performance of the program on the basis of them. Even so, the information generated by these outcome indicators is important to us, and to stakeholders who are interested in the performance of our program. Patterns in numbers of smokers, or incidence of smoking-related disease, are relevant to the design and implementation of the program, and well as to public perceptions about the relevance of the program. As responsible managers you have to demonstrate that you are monitoring this data, and feeding it into the ongoing operation of your program, even if you are not holding yourself accountable for the outcomes related to the data.

9.   Keep it Simple!

As the Office of the Auditor General has observed, unnecessary complexity is "one of the biggest risk factors"[5] that threatens the successful implementation of your performance framework.

When you look at performance frameworks that have been prepared for public programs you often see a tendency to want to include every detail of the program in the framework. You often see a very high number of performance indicators, and extremely complicated logic models. The result, as the Auditor General has observed, can be deadly:

> Over-complexity of a performance measurement system will lead to
> implementation problems and will simply frustrate stakeholders. The easier it is
> to use and apply, the more likely stakeholders will adopt and embrace the new
> approach. One way to keep it simple is to limit the number of indicators …
> (T)hree indicators which are solid measures of outcome are better than 10 which
> don't measure anything relevant.[6]

An overly complicated performance framework inevitably leads to implementation problems. Remember that your performance framework is not something that you create and then forget about. It is a "living document" that has to be managed and updated if it is to be of any use for management and accountability purposes. Each and every performance indicator within the framework needs to be serviced. The equation is simple: the more indicators you have, the bigger the administrative burden.

An overly complicated performance framework will also get in the way of telling a clear and credible performance story about your program. The bigger and more complicated your performance framework, the less likely it is that stakeholders will understand what it means. A performance indicator is supposed to be an expression of what is most *important* about the program. You measure a particular aspect of program performance because you regard it as highly significant in view of what your program aims to achieve. So if the framework is too complicated, it is probably because the people who built it

---

[5] "Implementing Results-Based Management. Lessons from the Literature," Ottawa: Office of the Auditor General of Canada, 2000, p. 7.
[6] *Idem.*

were, themselves, unsure about the kind of performance story they wanted to tell. They may not have had a clear picture of the high-level outcomes to which the program was supposed to contribute, and of the major steps by which the program was supposed to achieve its objectives.

How many performance indicators are too many? There are no hard and fast rules. In the interest of simplicity, think in terms of challenging every proposed performance indicator. Ask:

- do we really need to have this indicator?

- can we handle the administrative burden that the indicator creates?

- is the right kind of data readily available to service this indicator?

- is the proposed indicator linked to some strategically important element of our program?

A performance framework is supposed to be a means to an end, not an end in itself. It will only be worthwhile if it contributes to better program management, better decision-making, a better contribution by the program to ultimate outcomes, and stronger accountability to Canadians. On the other hand, complexity for complexity's sake contributes to nothing. It is a waste of time and effort!

10.   Beyond Good Performance Measurement

This Guide has tried to emphasize the point that good performance measurement is as much about a way of thinking as it is about a set of tools or techniques. It is a mindset that pushes you to think about ends rather than means. It compels you to think about outcomes that matter to Canadians, and about whether the logic of your program's design is suited to delivering those objectives.
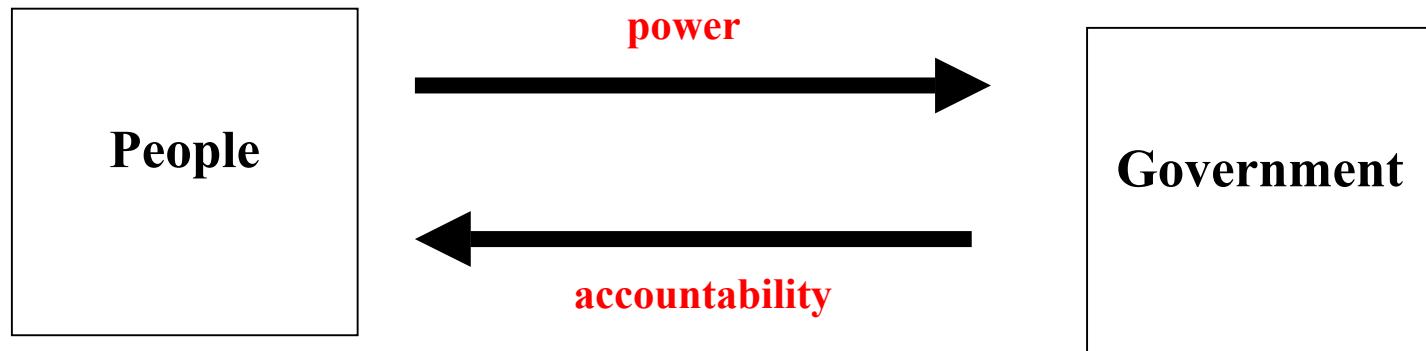
Whether you are measuring the performance of a public program, or whether you are designing or implementing the program, the basic principle is always the same: keep your eye on what you are trying to change out in the world that will, ultimately, make a difference that matters to people. Paying attention to this principle will increase your odds not only of developing a better performance framework for your program, but also of developing a better program – period.

This reminds us of the most basic point of all. Performance measurement and performance management are not ends in themselves. They are means to the end of better public programs that make a positive contribution to the lives of Canadians. Performance frameworks aren't worth the paper they are written on – and certainly not worth the trouble of producing them – if they don't contribute to that.

Figure 1

The Basic Bargain

Government

Figure 2

Obligation to explain and to justify how it has discharged its responsibilities

What did you accomplish?

What did you do?

If things went wrong, what did you do about it?

Why did you do it?

Are you doing what we want you to do?
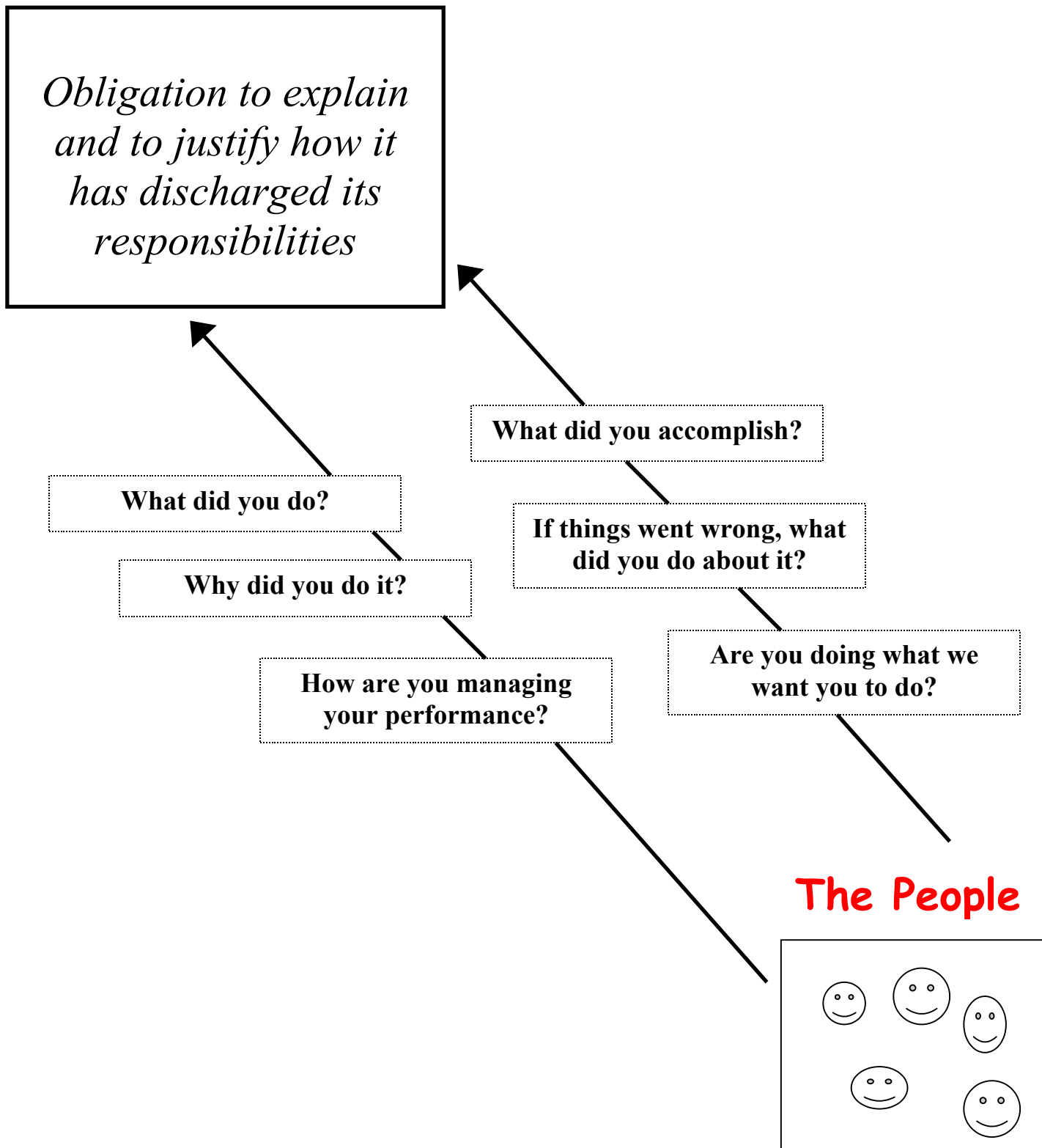
How are you managing your performance?

The People

Figure 3

33

Keeping Busy . . . Making a Difference . . .

Inputs

- resources applied

- *$ and FTEs*

Activities

- steps taken to carry out a program

Outputs

- program products

Outcomes

- changes out in the "real world"

- contributing to these is the reason for your program's existence

How We Keep Busy   *How We Make a Difference*

Figure 4

35

Generic Logic Model

Inputs

Activities

Outputs

Outcomes

immediate

intermediate

ultimate

Figure 5

37
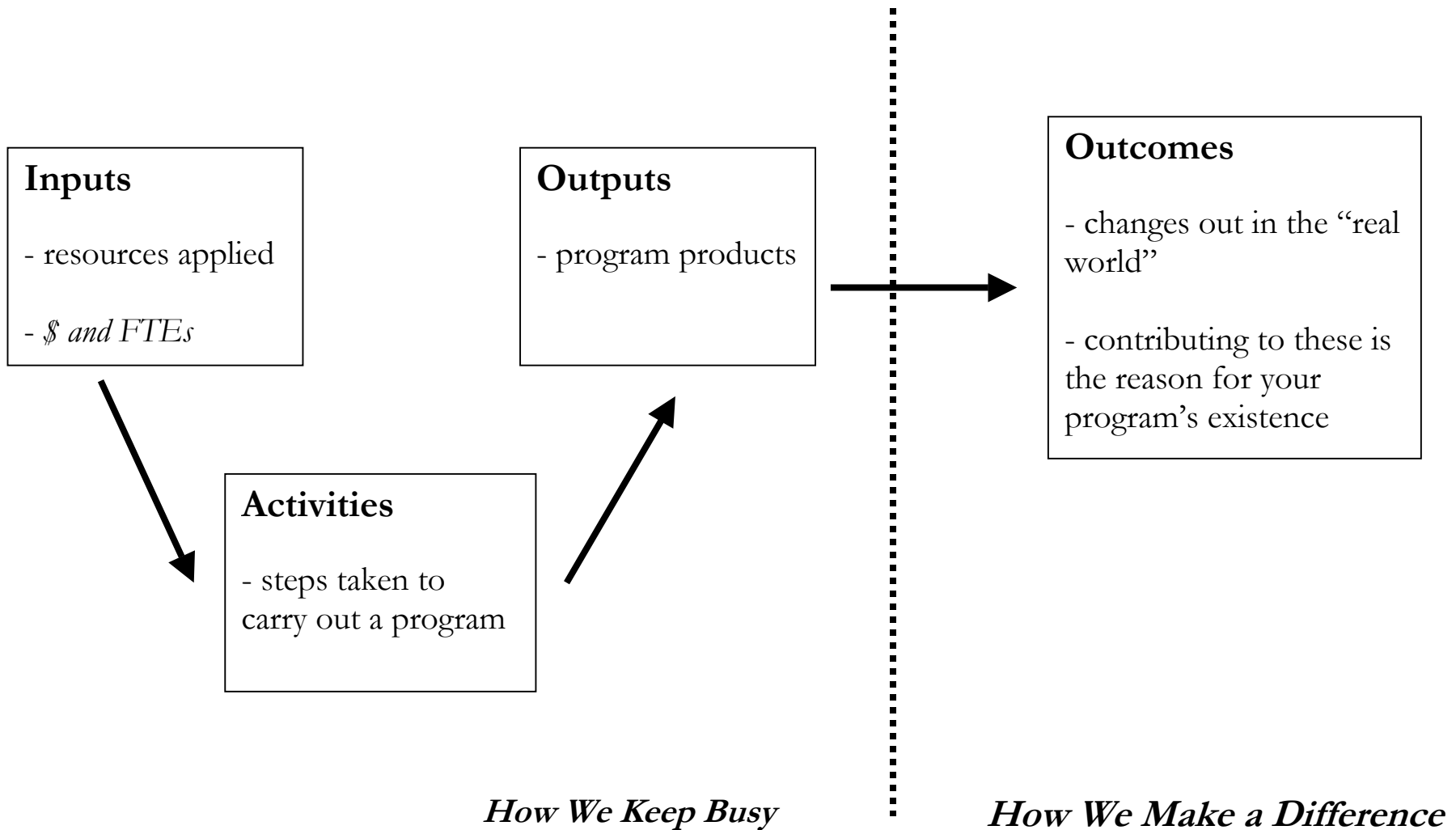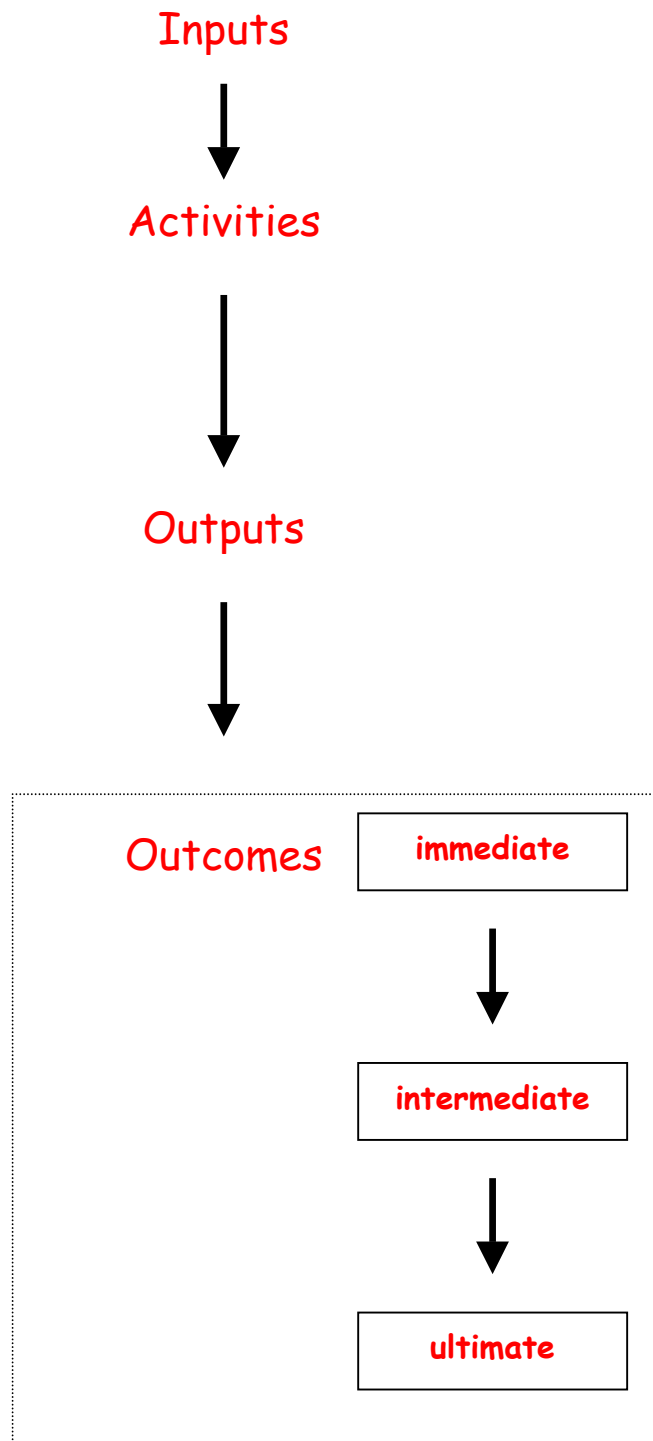
Example of a Logic Model
(anti-smoking TV ad campaign)

Inputs
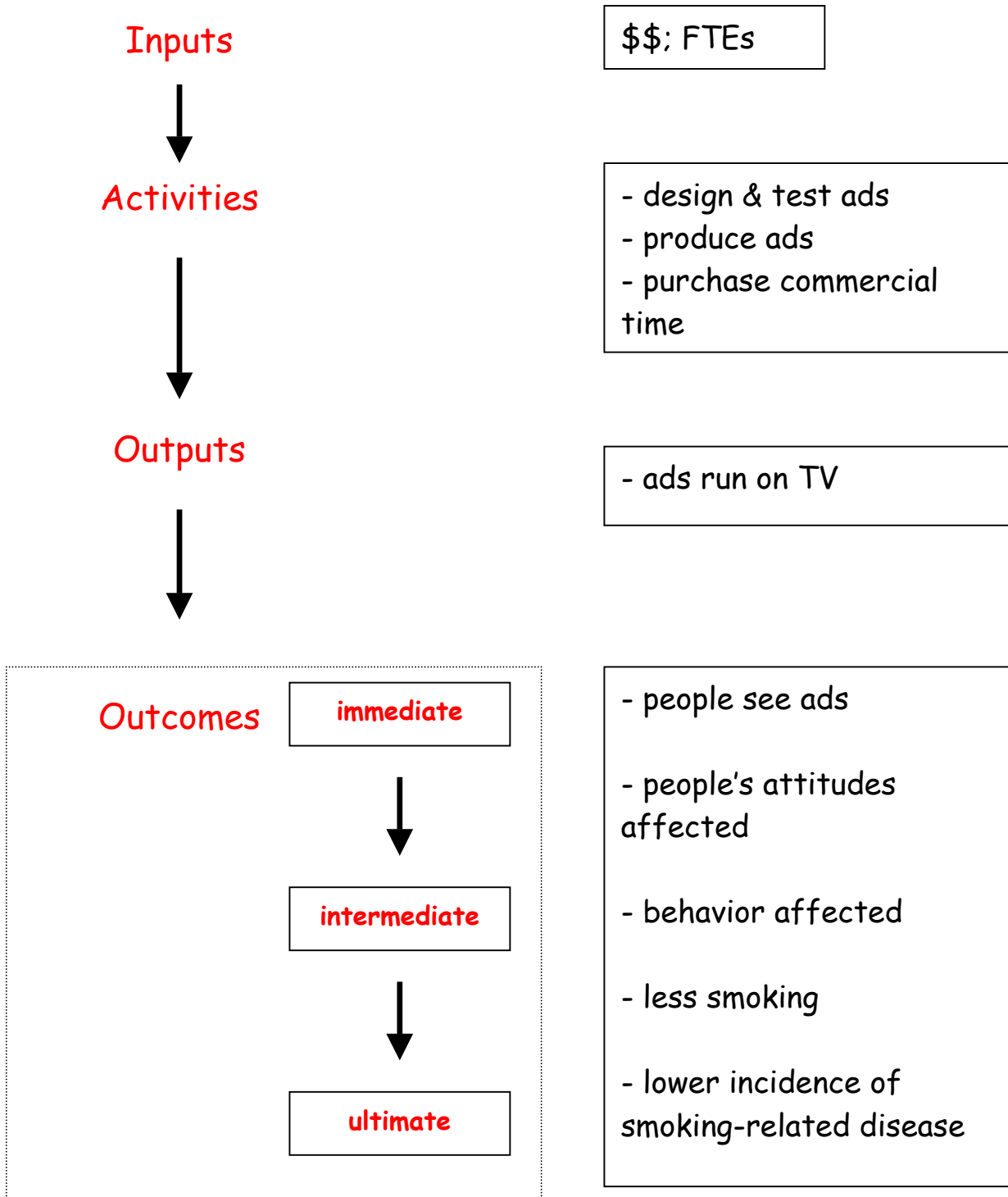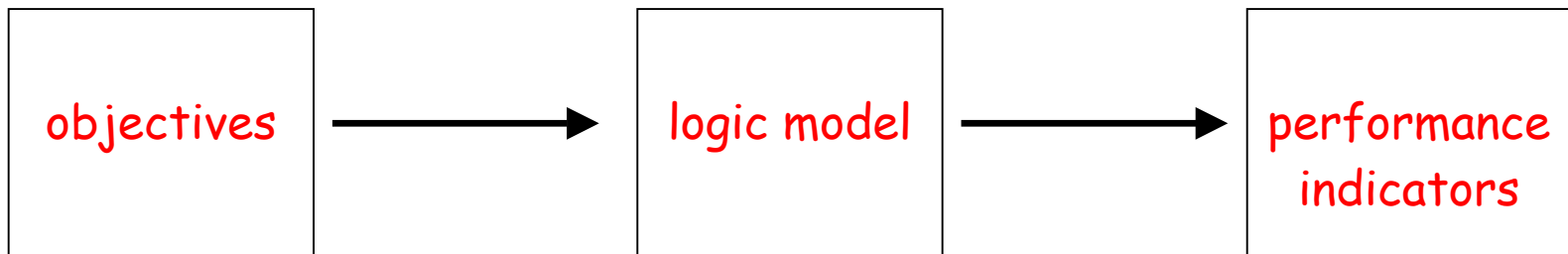
| $$; FTEs |

Activities

- design & test ads
- produce ads
- purchase commercial time

Outputs

- ads run on TV

Outcomes

immediate

intermediate

ultimate

- people see ads

- people's attitudes affected

- behavior affected

- less smoking

- lower incidence of smoking-related disease

Figure 6

Figure 8

41

Indicator Tradeoffs
Anti-smoking TV Campaign

Airtime for Ads

	L	M	H
Meaningful			
Difficulty			
Control			
Perverse Risk			
Annual Change			

Number of People Who Watch

	L	M	H
Meaningful			
Difficulty			
Control			
Perverse Risk			
Annual Change			

Indicator Tradeoffs (2)
Anti-smoking TV Campaign

Figure 9

Figure 10

Indicator Tradeoffs (3)
Anti-smoking TV Campaign

Less Smoking

Meaningful

L   M   H

Difficulty

L   M   H

Control

L   M   H

Perverse Risk

L   M   H

Annual Change

L   M   H

Smoking Related Disease

Meaningful

L   M   H

Difficulty

L   M   H

Control

L   M   H

Perverse Risk

L   M   H

Annual Change

L   M   H

Figure 11

Two Kinds of Indicators
(anti-smoking TV ad campaign)

Inputs

$$\$\$; \text{FTEs}$$
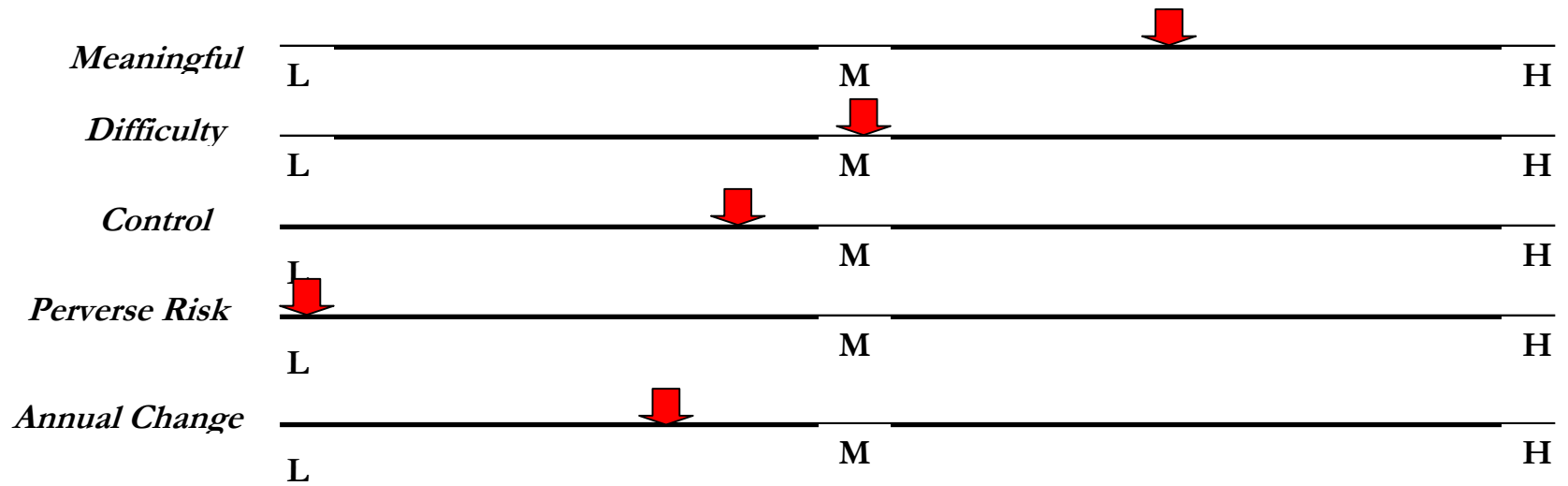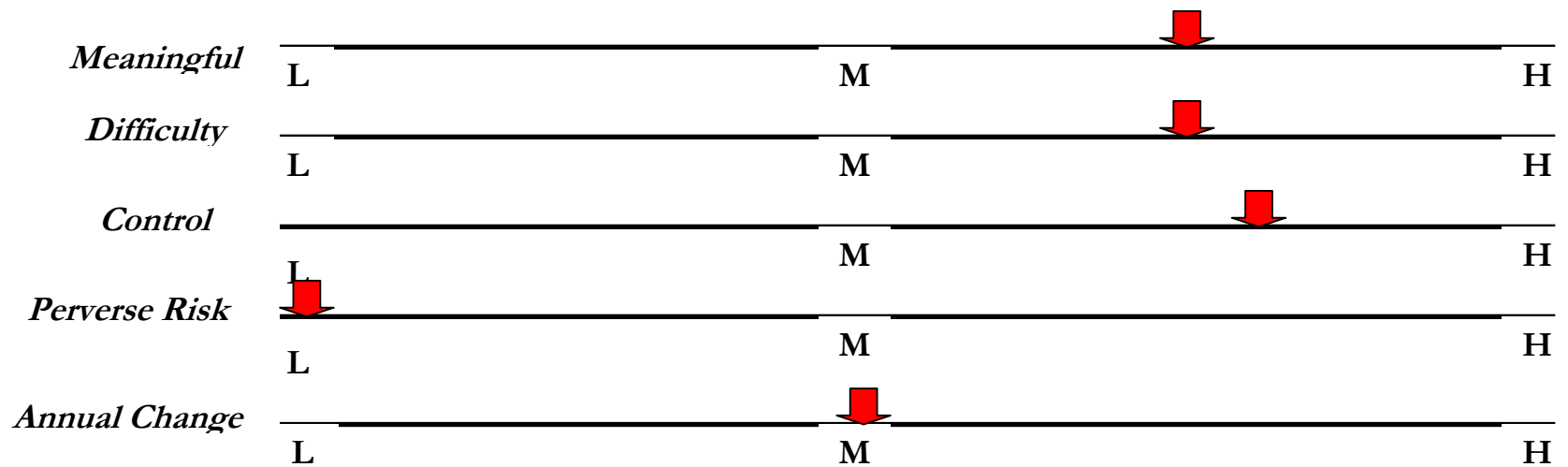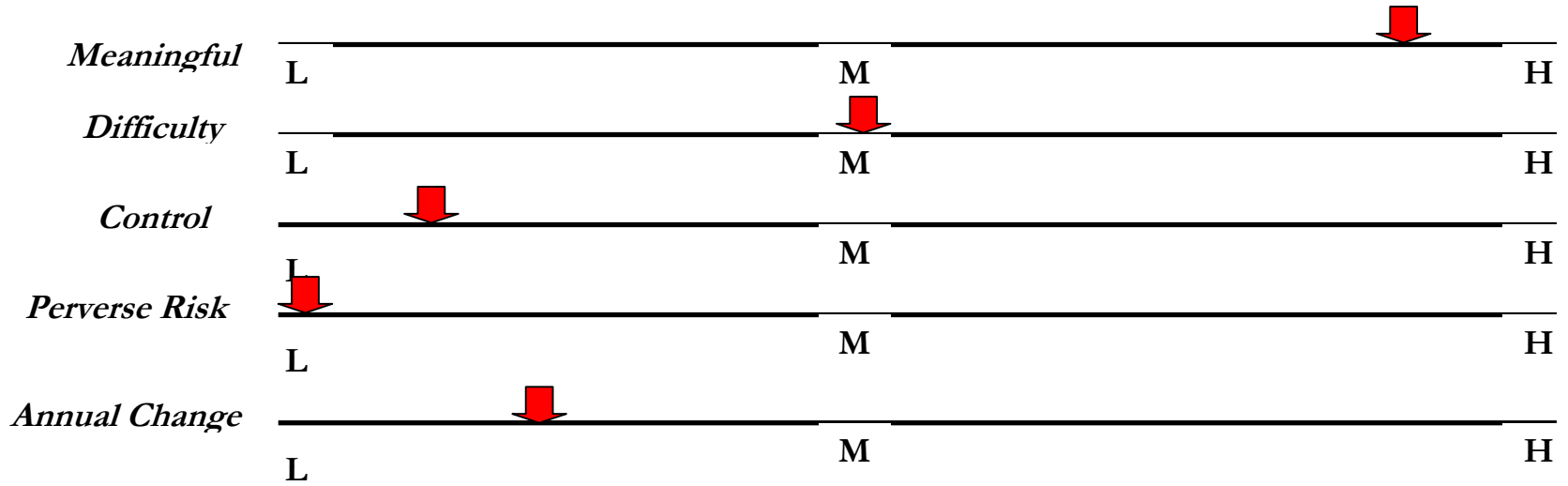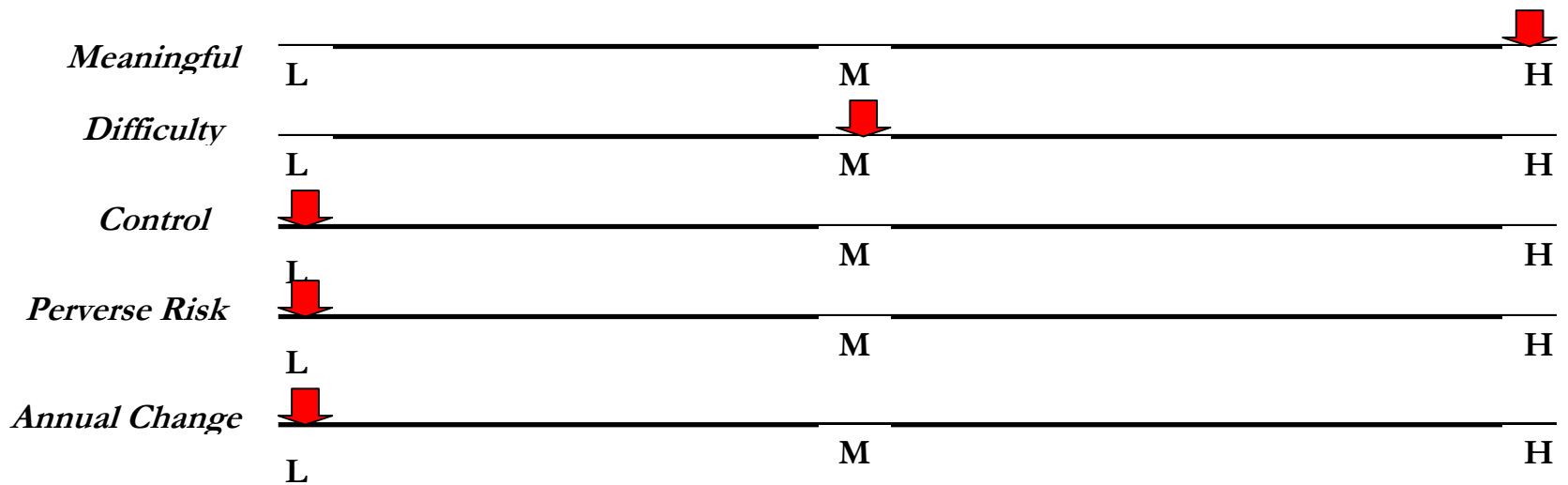
Activities

- design & test ads
- produce ads
- purchase commercial time

Outputs

- ads run on TV

Outcomes

immediate

intermediate

ultimate

- people see ads

- people's attitudes affected

- behavior affected

- less smoking

- lower incidence of smoking-related disease

Fair Reflection of Program Performance

Relevant to Program Performance
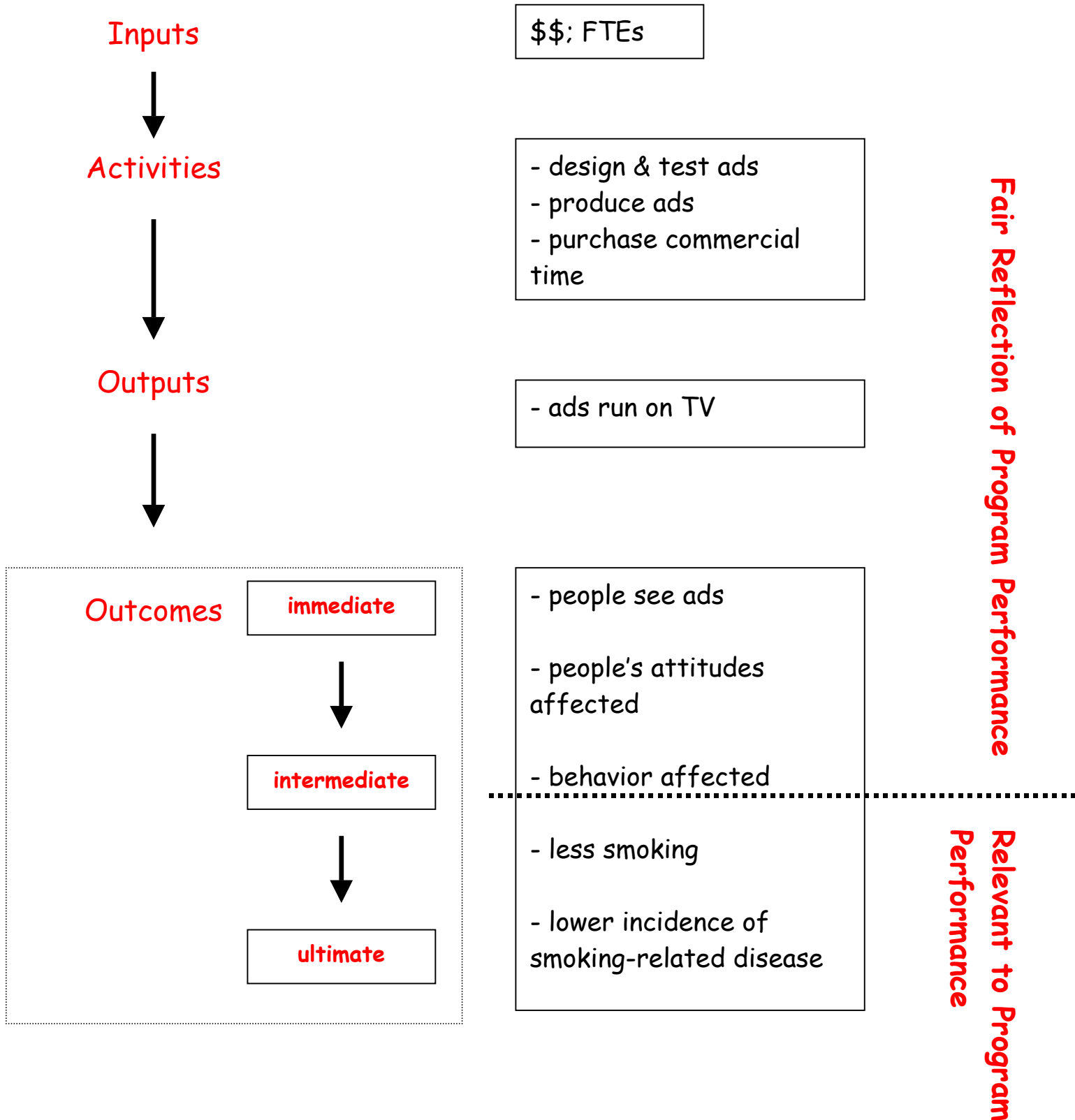
The Institute On Governance (IOG) is a non-profit organization founded in 1990 to promote effective governance. From our perspective, governance comprises the traditions, institutions and processes that determine how power is exercised, how citizens are given a voice, and how decisions are made on issues of public concern.

Our current activities fall within these broad themes: building policy capacity; Aboriginal governance; accountability and performance measurement; youth and governance; citizen participation; governance and the voluntary sector; and technology and governance.

In pursuing these themes, we work in Canada and internationally. We provide advice on governance matters to organizations in the public, private and non-profit sectors. We bring people together in a variety of settings, events and professional development activities to promote learning and dialogue on governance issues. We undertake policy-relevant research, and publish results in the form of policy briefs and research papers.

You will find additional information on our themes and current activities on our web site, at www.iog.ca.