



The Wisdom of Crowds – It Applies to Performance Measurement Too!

Steve Montague 2005

The viewpoints and brainpower of the many are almost always superior to the thinking and decision-making of the few.

James Surowiecki, in *The Wisdom of Crowds*, cites the example of 700 people estimating the weight of an ox, and coming on average within one pound of the actual weight. He extols the virtues of various 'markets' or crowds to successfully predict sporting events, set fair prices to assure product 'clearance' and even to focus appropriate blame for the air shuttle disaster – with only very partial information. This thesis flies in the face of much conventional wisdom which suggests that only a few key people should make judgements on various things – especially when matters are technical and complex.

To some extent we have a parallel situation in public performance planning and reporting. There has been a tendency to focus on a few key ideas and measures – without recognizing the need to have other indicators or measures to complement, confirm and contrast the perspective. Examples of overly selective measurement abound.

At the direct delivery level we have seen call centres and information services focus on speed – often to the detriment of service outcomes. In one case, an information booth was experiencing increased usage. Managers initially thought that this was due to its popularity – after all – they had just instituted a performance measurement 'standard' of service to all clients in two minutes or less. Unfortunately, most visitors to the booth, located in a large national park, were looking for directions. They were getting speedy – but insufficient, directions. They were going out, getting lost and then coming back to rejoin the line for more detailed information.

An early case of over-reliance on one performance measure was the use of benefit – cost analysis by the United States Tennessee Valley Authority. It turns out that ambitious engineers, with an

interest in building dams, were able to manipulate the ratios so that they would almost always come out to at least the threshold level to be able to build. Naturally, several of the harder to cost environmental concerns got short shrift in this approach, and the region is still paying the environmental ‘price’.

At the government wide level we have seen singular indicators used as a kind of rallying cry for new initiatives. Two recent Canadian federal examples come to mind – the notion that we should double the number of exporters, and the idea that Canadians should become the most connected people in the world.

These single-minded slogan-measures shared a similar conceptual problem, and each had specific technical problems. The conceptual problem they shared can be summarized by asking ‘Why?’. Why do we want to double the number of Canadian exporters? Why do we want to be the most ‘connected’ people in the world? What distinct need or gap was there in our situation that required these singular goals?

In the case of the exporters – it wasn’t at all clear that the raw number of exporters was a problem. In fact, given that evidence pointed to the reality that first time exporters often lost money in the pursuit of their new export sales, it might be counter-productive to encourage a large increase in first time exporters in a short period of time. (This was validated by the economic cycle which headed into a down turn in several key markets towards the late 90s and into the early 2000s. This kind of emphasis may have helped cause the collapse of many export oriented companies – particularly in the areas of high tech.)

Technically the ‘doubling the number of exporters’ goal also had a big problem – the Canadian Government didn’t have a good handle on the exact number of exporters that existed when the commitment was made – so it had a very difficult time figuring out when the number doubled.

As for the ‘being the most connected people in the world’ goal, the technical problem here was one of deciding what constitutes ‘connected’? The early efforts showed that, if you counted cable TV hook-ups, that Canada was close to number 1 in connectedness. But that wasn’t quite what most people had in mind. Expensive efforts to define a ‘connectedness index’ were undertaken.

The lesson in these recent cases is that we need to think carefully before picking single measures to represent the *raison d’être* of our new initiatives.

Now we face the threat of further over-simplification with regard to our health care system and other public initiatives. We have heard much discussion of the need to reduce waiting lists for health care procedures. While this goal is laudable, it would be dangerous to set it out in isolation. Just as with the liabilities noted in the case of a call centre focusing on speed over quality – a health care system which measures, and therefore emphasizes – speed over quality might not just be ineffective – it could be dangerous.

In the logical extreme, the emphasis could evolve to one of ‘people processing’ rather than health outcomes. An emphasis on speed over quality has led to death in the food services sector. (e.g. the fast food chain Jack-In-The-Box had a famous incident where service speed pressures led to

under-cooked hamburger which had lethal consequences.) One need not go far to envision an even more deadly scenario in hospitals and health care units. Indeed some may argue that this scenario has started to play out already – as we see staff infection rates, ‘sick hospital’ syndromes and re-admission rates rise in certain areas – we have to wonder about a system which – in Ontario at least – put out a scorecard for its hospitals in 2001 which had indicators for resource use, speed and efficiency significantly outweighing indicators of health outcomes. (See Hospital Report 2001 – The Ontario Hospital insert in The Globe and Mail, July 27th, 2001.)

So what is the answer? Surely we aren’t advocating dozens of measures for every goal? (This is probably the # 1 concern expressed by our clients – the fear that as performance measurement specialists and consultants we will build a measurement scheme which will require a burdensome bureaucracy of its own to administer.) Well, we’re not advocating dozens of measures – but we *are* suggesting that a good set of measures form a small crowd or cluster which do the following:

Complement: It is useful for measures or indicators to complement other measures. Complementary measures provide greater insight into a key concept. As an example, many groups will use a quantitative measure for items such as reach, take-up, usage or client satisfaction. A complementary measure to the quantitative total is often a breakdown of, for example, usage by key target groups. The ‘spread’ or mix of users can tell you as much or more about the appeal or value of your service as the total # of users.

Confirm: A good measurement system allows for the confirmation that an indicator truly measures what you think it measures. An example of this occurred when we gathered complementary qualitative information on satisfaction. We asked the question ‘Why?’ after getting people’s satisfaction rating on service. It turned out that people rating themselves as ‘somewhat satisfied’ were often really not satisfied when one analyzed their qualitative comments. This allowed us to adjust our interpretation of satisfaction scores.

In a second, more recent example, complementary measures might have helped pollsters to more accurately predict the outcome of the 2004 Canadian federal election. Rather than tallying the numbers from the simple response to “If the election was held today who would you vote for?” – several analysts, with the benefit of 20/20 hindsight, pointed to people’s attitude towards the different leaders, their prediction of (as opposed to their *preference for*) the likely winner, as well as the ‘firmness’ of their preference – as measures which didn’t confirm the straight polling data.

Contrast: The contrasting measure is the balancing measure. The most well known example of this phenomenon is the emergence of the Balanced Scorecard over the last decade. The Balanced Scorecard evolved as a response to narrowness in measurement – perceived by Harvard professor Robert Kaplan and his colleague David Norton. In their early Harvard Business Review articles and their first Balanced Scorecard book, they suggested that companies in the early 1990s focussed too strongly on financial indicators as performance measures. This was seen to be akin to “driving by looking out the rear view mirror”.

A good Balanced Scorecard would thus have an appropriate mix of outcome measures (“lagging indicators”) and measures of the drivers of future performance (“leading indicators”). Their

argument for balance goes like this. Lagging indicators without leading indicators do not communicate how the outcomes are to be achieved, nor do they provide an early indication about whether an organization's strategy is being implemented successfully. Conversely, leading indicators without lagging indicators may point to the achievement of short term operational improvements but will fail to reveal whether these operational improvements have been translated into meaningful medium term outcomes and, eventually, to desired final outcomes.

In essence, a balanced set of measures can help organizations to *optimize* rather than *maximize* key aspects of performance. In this way the Balanced Scorecard, like the 'results logic' approach familiar to many public sector analysts, is not merely a collection of leading and lagging indicators. It is the translation of the organization's strategy into a linked set of measures that define both the long-term strategic objectives, as well as the mechanisms for achieving those objectives.

In summary, a good measurement system will recognize that single indicators representing the ideas and concepts of performance are not just annoying – they can be downright misleading and dangerous. One way to combat this is to recognize the wisdom of crowds or clusters. This applies to clusters of *indicators* which complement, confirm and contrast, crowds or clusters of *tests* which independently replicate findings – and crowds of different *stakeholders*. Indicators should be drawn from the behaviours of different groups – different segments of users (including non-users), different geographic regions, different levels, cultures etc.

The *real* conclusion – consistent with Surowiecki's message – is that the best measurement 'system' – like the best decision-making system – is simply one that involves diverse, freely exercised feedback. Our public measurement frameworks need to preserve this value.

Steve Montague is a founding partner of Performance Management Network Inc. and a founding member of the Performance and Planning Exchange (PPX) – a not-for-profit organization with members from the public sector, private sector and academia, dedicated to improving results and performance management through the exchange of information and ideas (see ppx.ca). Steve provides consulting and education services to clients in the Canadian federal, provincial, regional and international government communities.

steve.montague@pmn.net