Performance Measurement: Issues, Approaches, and Opportunities

Nicholas Fisher[1,2]

**[1]School of Mathematics and Statistics, University of Sydney, New South Wales, Australia,
[2]ValueMetrics Australia, North Sydney, North South Wales, Australia**

ABSTRACT

Performance measures permeate our lives, whether or not we are aware of them. They can support or frustrate what we are trying to do, help or hinder enterprises going about their business, encourage or distort behaviors, clarify or confuse purpose. We illustrate some of the consequences of poor performance measurement, explore some of the reasons why poor metrics are in use, and describe a systematic way to look for performance measures in a variety of settings. There are real opportunities and challenges awaiting an inquiring and creative data scientist.

Keywords: due diligence, Performance Measurement Framework, quality management, stakeholder value management

1. Introduction

Many aspects of our working lives—let alone our private lives—are subject to ongoing assessment, and sometimes, measurement. Whether we function as an individual, as a business, as a university, as a government, someone else is asking questions about our performance. Are we going to meet our monthly target for new sales? Are our workforce safety statistics improving? What do we expect our research grant income to be? How can we demonstrate to the community that our efforts to reduce carbon emissions are working?

However, each of these questions is forward-looking. Many more queries relate to measurement of what has actually occurred, when the gaps between promise and reality are cruelly exposed. It is therefore clearly of interest to individuals, enterprises, and governments to have sensible (quantitative) targets, and sound ways of assessing progress toward these targets.

The operative terms in the previous paragraph are "sensible targets" and "sound ways." On what basis can we select them? There is clearly scope for improvement: we have all been subject to the imposition of seemingly arbitrary numerical targets; some of us have actually set such targets. Muller (2018) has dedicated a whole book to what he terms "the tyranny of metrics." And such targets have real consequences: as Eliyahu Goldratt (1990, p26) observed, "You tell me how you measure me, and I will tell you how I will behave."

Building upon early work (e.g., Dransfield et al., 1999; Fisher, 2013; Kordupleski, 2003), this article aims to stimulate data scientists to engage with the general area of performance measurement by developing better approaches and pursuing their adoption by government, academe, business and industry, and indeed in all areas of human endeavor, because whatever people are doing, performance measurements are lurking to make judgments. Good performance measurement is increasingly at the 'pointy' end of good management. In a world becoming more and more data rich, leaders and managers will become more reliant on data scientists to assist them in understanding and interpreting what is going on in their business, and an increasingly important

component of this will relate to how they measure all aspects of performance, and then decide on consequent actions.

Specifically, Section 2 seeks to illustrate the wide range of circumstances in which problems have arisen as a consequence of poor performance measurement. Some of the reasons these problems have arisen are explored in Section 3. Section 4 describes an approach that has been developed to tackle performance measurement issues in a number of settings, and Section 5 contains some thoughts on opportunities to be explored and how to get started in the area.

2. Examples of Why Performance Measurement Matters

The following examples span the spectrum from simple performance indicators that distort the behavior of individuals to whole suites of measures that fail to assist members of a board in doing their jobs properly. They will raise points that anticipate the content of the next section.

The original motivation for this research campaign was a desire to know what sorts of metrics were needed to manage a research group. The natural starting point was to ask the question: 'How do leadership teams decide which metrics to monitor?' There was little or no guidance to be found in the literature at that time (early 1990s). Of course, boards of companies routinely monitored a large number of financial indicators and statutory reporting numbers, but there was very little use of non-financial indicators. This *appears to be* consistent with a management philosophy that the purpose of a company was to make a profit for its shareholders (a philosophy that is now losing credibility) and that non-financial metrics have little role to play in this.

Remark. This last observation relates to so-called 'dry' economic theory. Zingales (2020) commented:

> The debate between a shareholder-centric perspective and a stakeholder-centric one dates back at least to the early 1930s, when Adolf Berle and E. Merrick Dodd argued these two positions in the *Harvard Law Review*. Yet, a crucial milestone in this debate has been set by Milton Friedman who, fifty years ago almost to the day, wrote in the *New York Times Magazine* that the only social responsibility of business is to increase its profits. Whether you love or hate his arguments, Friedman has set the terms of debate for the last 50 years.
>
> Since the mid-1980s to the early 2000s, Friedman's position was dominant not only in academia, but also in the business world. In 1997, the Business Roundtable proclaimed that "the principal objective of a business enterprise is to generate economic returns to its owners," all but endorsing Friedman's position.

An early counterexample to this traditional view is provided by the AT&T story which, although dated, we introduce because it resulted in the development of a process critical to the general approach that we have developed.

2.1. The AT&T Crisis

In 1986 AT&T was confronted by a business paradox. As a company of some 300,000 employees, operating in 67 market sectors in 32 countries, they were surveying 60,000 customers a month to ascertain their satisfaction with AT&T's offerings. The overall customer satisfaction rating was 95%. However, at the same time, they lost 6% market share, where 1% was worth $600,000,000. For the first time in corporate history, AT&T laid people off—25,000 worldwide. The problem turned out to relate to the definition and measurement of customer satisfaction, and led to a revolution in how market research was conducted (see Kordupleski, 2003, for a full account.) There was a subsequent dramatic and sustained improvement in AT&T's financial performance, the key customer metric became a critical item in AT&T's quarterly board reports, and a component of individual remuneration packages for senior officers was linked to this metric. We shall return to this story in Section 4.1.

2.2. The Hastie Group Collapse

The Hastie Group, a multinational organization supplying a wide range of mechanical, electrical, hydraulics, and refrigeration services to the building and infrastructure sector, collapsed in 2012, owing over a billion dollars. The company PPB Advisory was appointed to wind up the Group. Its report identified serious deficiencies with the overall control of the Hastie Group, including:

- internal systems for project management were inadequate and not to industry standard;
- financial reporting from subsidiary level up to group level was not uniform and open to manipulation; and
- the board of Hastie Group did not appear to "adequately challenge divisional/subsidiary results or forecasts."

In other words, the Group did not know what was going on because they lacked a good Performance Measurement System and, in particular, regular board reports that answered—or at least, strongly invited—the questions *Where are we now*? *Where are we heading?* and *Where do we need to focus attention?*

As an extension of this conclusion, we contend that a good Performance Measurement System should provide the following:

1. A concise overview of the health of the enterprise.
2. A quantitative basis for selecting improvement priorities.
3. Alignment of the efforts of the people with the mission of the enterprise.

2.3. The Sunbeam Bankruptcy

For a number of years, Al Dunlap was the darling of Wall Street for his seeming ability to enhance shareholder value in companies he took over, by closing factories, firing staff, and pursuing very aggressive quarterly sales targets. See Byrne (1999) for the full story of Dunlap's career. In 1996, he was appointed chairman and CEO of Sunbeam, with a remuneration package structured in such a way that he benefited whenever Sunbeam met quarterly sales targets, with subsequent increases in Sunbeam's stock price. As it proved impossible to meet

these targets by principled means, Sunbeam resorted to the dubious business practice of so-called bill-and-hold strategies. Thus, in the winter of 1996, Sunbeam recorded record sales for gas barbecue grills, having persuaded retailers to take advantage of significant discounts on appliances they would not be selling to consumers for another 6 months. Sunbeam billed these as sales, while having to hold the appliances in warehouses. Having brought these sales forward, things were even worse in the following summer. Eventually, all the creative accounting plus a host of other factors led to Sunbeam filing for bankruptcy and Dunlap being fired and disgraced.

Examples like this one arise to this day. However, the story of Sunbeam's descent into bankruptcy is notable for its sheer size.

We conclude this list of board-level examples with a story that has a happy ending born out of a bad experience.

2.4. Motorola and Six Sigma

It is well known that the Six Sigma methodology for continuous improvement was originally developed by Motorola. What is less well known is how it came about. As recounted by Debby King-Rowley, formerly global director of executive education at Motorola,

> There was a very early focus on cycle time reduction across the board. That was introduced from the C-suite[1] down in 1986. At that time, Motorola was working on quality through the 3 leading guru schools of thought at the time—Deming, Phil Crosby, and Juran. No single approach was being promoted from corporate. When cycle time was introduced, it was introduced as part of a 3-legged stool—cycle time, quality, cost. All three had to be in balance within the business units. Cycle time was the only one being driven (in goal of 50% reduction and process) from headquarters. Once cycle time focus was in place, an eye was turned to quality to standardise the approach on a company-wide basis. A lot of work was being done with Deming's concepts, but an internal electrical engineer, Bill Smith, in our then "Communications Sector" created the concept of Six Sigma. He took the idea to Bob Galvin, who is quoted as telling Bill "I don't fully understand it, but it seems to make sense. Come meet with me weekly 'til I understand it." Bill did, Bob fully grasped it, then others (particularly statisticians like Mikel Harry) were brought in to support and advance Bill's Six Sigma. Then the rest is history.
>
> (Email communication from Debby King-Rowley to the author, September 13, 2008. Reproduced by permission.)

However, the cycle time metric was introduced first, so that everyone focused on reducing how long each process took to complete. The totally predictable outcome was a blowout in waste and rework because process capability (the ability of the process to produce outputs conforming with specifications) was being overlooked. Adding the six-sigma component addressed this issue and so helped control costs. The well-packaged Six Sigma methodology has since been very beneficial for many enterprises.

Examples relating to board reporting might seem far removed from the daily lives of most of us. However, there is one particular marketing metric that not only occurs in board reporting but that pervades our everyday interactions with enterprises great and small.

2.5. Net Promoter Score

As will be explained later in the article, a lot of market research and resources are needed to produce the key customer metric developed by AT&T. However, the result is a robust market research process of continuous improvement, together with a suite of metrics that provide people at all levels in the enterprise with the information they need to do serve customers well, quite apart from enabling an enterprise to operate competitively.

In contrast to AT&T's approach, Reichheld (2003) asserted that there was no need for expensive market research campaigns, and all that mattered as far as customer satisfaction was concerned was a single number— Net Promoter Score (NPS) —defined along the following lines:

> After an interaction with a company's products or services, people are asked "How likely is it that you would recommend our company to a friend or colleague?" Based on their responses on a 0 to 10 rating scale, group the respondents into "promoters" (9–10 rating—extremely likely to recommend), "passively satisfied" (7–8 rating), and "detractors" (0–6 rating—extremely unlikely to recommend). Then subtract the percentage of detractors from the percentage of promoters.

NPS has been very influential. It has been adopted by some of the world's largest enterprises, both public and private, with a view to monitoring customer satisfaction and at the same time slashing marketing budgets. Unfortunately, what has also been slashed is the capability of an enterprise to identify how it is performing and where it needs to focus improvement priorities. This was discussed in Fisher & Kordupleski (2019), and some of the problems are summarized in Table 1 (derived from table 4 of Fisher & Kordupleski, 2019). There is little evidence in the form of case studies of any sustained benefits that enterprises have derived by adopting NPS to manage their responses to market needs, let alone benefits to customers to compensate for the nuisance requests to provide NPS ratings.

Remark. Net Promoter Score (NPS) has also been (trivially) adapted for use with staff satisfaction surveys, specifically, to measure 'employee engagement.' Unsurprisingly, it performs poorly in this setting as well, for similar reasons.

Table 1. Evaluation of NPS against some desiderata for satisfaction surveys.

Desiderata for (customer) satisfaction research	NPS response

a. A statistically sound method that ensures that no important attributes of the product or services have been omitted from the survey.	Totally ignored. The current routine collection of NPS scores for operational activities (cf. the introductory example in Section 2) is purely observational, with little or no understanding of demographic factors let alone sampling biases.
b. A means of linking survey results to higher-level business drivers, and an empirically proven lead indicator of business improvement (for example, in return on investment).	Assertion of "correlation with growth and profitability."
c. Actionable survey results, including the ability to drill down.	No.
d. A means of identifying where to focus improvement priorities so as to have the greatest beneficial impact on the business bottom line.	No defensible method.
e. Meaningful benchmarking metrics.	Nothing below level of NPS; nor NPS itself. Also, there are no agreed standards about how NPS should be aggregated in a company to produce an all-of-company metric.

2.6. Three Bibliometrics That Can Drive Undesirable Behaviors

The academic world provides many performance measurement challenges. There is an ongoing and urgent need to be able to evaluate the quality of research, or to compare individuals (promotion cases, appointments), departments, graduate schools or universities, or to rank applications for research funding, or to be able to assess teaching quality or student course experience. Indeed, it has produced its own specialized set of performance measures known as 'Bibliometrics' in an attempt to quantify the notion of research quality.

Here we list three most commonly used metrics; see Adler et al. (2009) for a careful discussion of these and others.

(a) Paper count. How many articles did you publish (last year / last 5 years)? This encourages people to report their research in terms of the smallest publishable segments, discouraging longer and more comprehensive treatments.

(b) Citation count. How many times has your research been cited by others?

(c) Journal impact factor. How many articles in a particular issue of the journal are cited by other articles not in that issue? One consequence of this (from personal experience) was strong encouragement from a journal

editor, after accepting an article for publication, to search through past issues of that journal for other articles that could be cited.

Clearly, these three metrics purport to provide some evidence of 'quality' —of the individual or of the journal. Whether or not this is the case, they raise a very important issue: *Who is the customer here?* After all, it is the customer who decides, or should be deciding, what 'quality' means for them. And this question leads to two essential and complementary elements to developing what I call a Performance Measurement System: *Stakeholder Analysis* and the *Tribus Paradigm*.

In these particular cases, conducting *Stakeholder Analysis* entails answering the question: 'Who are the people or groups with a vested interest in how well this work is done, and what does "quality" mean to them?' This leads us to an important line of thinking due to Myron Tribus, the Tribus Paradigm (although not styled as such, when communicated to the author by email, around 1993):

Step 1. What products or services are produced and for whom?

Step 2. How will 'quality,' or 'excellence' of the product or service be assessed and how can this be measured?

Step 3. Which processes produce these products and services?

Step 4. What has to be measured to forecast whether a satisfactory level of quality or excellence will be attained?

This paradigm captures many critical elements:

1. The starting point for selecting performance measures, whether for an individual or an enterprise, is the customer; hence the need for stakeholder analysis.
2. For each of the answers to (1), the concept of 'quality' has to be identified. (1) and (2) then form the basis for measuring a good *outcome* for the customer.
3. A fundamental precept of good management is that delivering good quality is achieved by process improvement.
4. Finally, one wants confidence that a good outcome will be achieved. This will come from identifying good lead indicators, or predictors, of the outcome metrics.

Remark. Myron Tribus was an American engineer, inventor, bureaucrat, management expert, historian, scholar and educator and life-long learner, and also a member of the American Statistical Association, and who published a well-cited book about decision theory. He knew W. Edwards Deming well, and found practical language and actions to explain and implement Deming's theory. Tribus devised the basic criteria that now underpin the Baldrige Awards and other similar Business Excellence frameworks. Fisher & Vogel (2017) provided an appreciation of his life and work (from an Australian perspective: Tribus's work had impact well beyond the shores of the United States).

2.7. Confusion in Measuring Research Outcomes and Impact

The Australian Research Council has attempted to make progress with the issue of measuring *research impact*; see https://www.arc.gov.au/policies-strategies/strategy/research-impact-principles-framework. This has led them to produce the information shown in Table 2.

Table 2. The Australian Research Council's basis for measuring impact of research.

Inputs	Activities	Outputs	Outcomes	Benefits
• Research income • Staff • Background Intellectual Property (IP) • Infrastructure • Collections	• Research work and training • Workshop / conference organizing • Facility use • Membership of learned societies and academies • Community and Stakeholder Engagement	• Publications, including e-publications • Additions to National Collections • New Intellectual Property (IP): patents and inventions • Policy briefings	• Commercial products, licenses and revenue • New companies – Spin offs, start-ups, or joint ventures • Job creation • Implementation of programs and policy • Citations • Integration into policy	• Economic, health, social, cultural, environmental, national security, quality of life, public policy, or services • Higher quality workforce • Job creation • Risk reduction in decision-making

Foreshadowing a discussion later in this article, it is appropriate to ask: What does impact mean, as far as measuring it is concerned? In my view—and as is implied by the Tribus paradigm—*'Impact' is the subjective assessment made by the customer for whom these outcomes are produced*. Viewed in this light, we need to ascertain the views of quite disparate customers about the various outcomes listed in the table. As things stand, they are activities—but whether they are or were good or bad activities is up to the judgment of the people for whom they were produced, and these judgments are not listed in the table.

For example, a simple version of this issue occurs when, say, a regional council lists, as a good outcome for their community, 150 community consultations in the last 12 months. The unanswered question is: Were they of any real value to the community?

And what about the impact of performance measurement on our everyday working lives? If you work for an enterprise, what sorts of metrics do you need to monitor, in order to do your job well? And what sorts of metrics should you report?

To respond meaningfully, you need to be able to answer three more seemingly simple questions:

Q1. What is your job?

Q2. What does it mean to do the job well?

Q3. How do you know you are on track to do a good job?

As we shall see in the next section, the issues raised by these questions go to the heart of performance measurement, most particularly to the meanings of—and critical difference between—*Accountability* and *Responsibility*. For those readers who have trouble finding satisfactory responses and who have direct reports, you can expect that your reports will have comparable difficulties.

3. Probing Causes and Consequences of Performance Measurement Problems

How did all the problems highlighted in the previous section come about? Clearly, company collapses (cf. 2.1, 2.2), the wastage of large amounts of money on misleading market research (2.5), and misdirection of scarce research funding (2.7) are circumstances best avoided (or at least minimized) if possible. On the face of it, one would think that many of these things would have been avoided by exercising plain common sense. The French philosopher Antoine Arnauld anticipated this issue a few centuries ago (Arnauld, 1662, p. 10): "*Le sens commun n'est pas si commun que l'on pense"* [Common sense is not as common as one thinks].

In this section, we explore some of the reasons why performance measurements are being chosen and some likely consequences.

3.1. Benchmarking

As defined in Wikipedia, "Benchmarking is the practice of comparing business processes and performance metrics to industry bests and best practices from other companies." Large enterprises also use benchmarking internally, to monitor the comparative performance of different business units: for example, by comparing the overall metrics for staff satisfaction surveys. Thus, it is an important tool in seeking to improve efficiency and maintain a competitive position.

However, it is not uncommon to find that the use of benchmarking is a false economy. One striking example was highlighted by the Australian Royal Commission into Financial Misconduct in the Financial Services Sector, where it emerged that Australia's largest banks were using Net Promoter Score as their ultimate customer satisfaction metric, rather than investing in proper market research.

Another failure of benchmarking is when an inferior survey instrument is used because 'everyone else in this industry uses this survey,' or because 'we want to be able to see how we're going now, compared with a year ago.' We elaborate on this under the next heading: Compliance.

3.2. Compliance

In many sectors, especially government and not-for-profit areas subject to a range of statutory requirements, Compliance in the guise of Quality Assurance appears to be the enemy of a culture of continuous improvement. For example, it may be a sector requirement that a safety culture survey be conducted every year or so. The easy—and lazy—thing to do is to see what others in the sector are doing and do the same thing. Maybe some actions come out of this, maybe not. A box has been ticked: we've done our annual culture survey. The numbers look roughly the same as the sector average or last year's numbers so we're okay for another 12 months. Considerations of whether the survey instrument actually conforms with important desiderata for such surveys, beyond demonstrating compliance and possibly providing some form of benchmarking, are often ignored. One very widespread example is based on variants of the Safety Awareness Questionnaire (SAQ) (see Fisher et al., in press, for an explanation of the significant deficiencies in this instrument and the multitude of survey instruments that it has spawned).

3.3. Metrics Not Linked to Process or Purpose

How often have you complied with a request to complete some sort of satisfaction survey—staff or customer or community or whatever—and then heard nothing further? Then, a year or two later, a repeat request arrives. People become cynical, and either refuse to comply (many staff surveys have very low response rates) or vent their frustrations, which will be ignored for another year or two.

Also, surveys of this type need to be more frequent than once a year. Any CEO who was told that financial updates would be provided once a year would rightly complain: 'How can I be expected to run this place when I never see what's going on? All I can do is react after the event.' The same applies to customer issues, staff issues, and what's going on in the way your enterprise interacts with the wider community. Information needs to be provided in a timely fashion, so that you can anticipate rather than simply react. One of W. Edwards Deming's more insightful aphorisms was: "Management is Prediction" (Deming, 1994).

One remedy for this is known as Price's Dictum (Price, 1984):

- No inspection or measurement without proper recording.
- No recording without analysis.
- No analysis without action.

In other words, the measurement must lead into a process (of improvement).

3.4. Failure to Take a Systems View

Goldratt's (1990) whole book is dedicated to this issue. What is supposedly optimal locally turns out to be suboptimal when viewed in the context of a bigger system. The COVID-19 epidemic provides an interesting example. Siddhartha Mukherjee, an Indian-American physician, biologist, and oncologist, published an article

in the context of a bigger system. The COVID-19 epidemic provides an interesting example. Siddhartha Mukherjee, an Indian-American physician, biologist, and oncologist, published an [article](#) in *The New Yorker* titled "What the Coronavirus Crisis Reveals About American Medicine" (2020). The article commences with a cautionary case study highlighting the limitation of a *[Just in Time](#)* (JIT) approach to production, a key aspect of which is to avoid the various forms of waste associated with stockpiling materials and components by having available only what is needed for immediate use. While this might be 'optimal' locally, within the confines of the production system itself, it lacks robustness against possibly buffeting within a larger system. Mukherjee's story relates to one of Toyota's component suppliers:

> At 4:18 a.m. on February 1, 1997, a fire broke out in the Aisin Seiki company's Factory No. 1, in Kariya, a hundred and sixty miles southwest of Tokyo. Soon, flames had engulfed the plant and incinerated the production line that made a part called a P-valve—a device used in vehicles to modulate brake pressure and prevent skidding. The valve was small and cheap—about the size of a fist, and roughly ten dollars apiece—but indispensable. The Aisin factory normally produced almost thirty-three thousand valves a day, and was, at the time, the exclusive supplier of the part for the Toyota Motor Corporation.

With JIT, the focus is on cycle times (responsiveness of suppliers, responsiveness to customers) and cost savings (little idle inventory or storage space). However, Mukherjee recounts how Toyota was able to recover within days because the greater system within which it operated—the whole Japanese manufacturing system! —had the capacity for ad hoc adaptation to supply the shortfall in components within a very short period of time.

Mukherjee contrasted this with the overall failure of the U.S. health care system to cope with the extraordinary demands placed on it by the COVID-19 pandemic. In fact, Mukherjee portrays U.S. health care management as having little or no ability to respond to a crisis. This is not so much a reflection on the many people working as best they can as of a dysfunctional system choked with obsolete compliance requirements and contorted by inappropriate performance measures.

Remark. In fact, Sommer (2009, p. 73) states that:

> Calling U.S. medicine either "health care" or a "system" is an exaggeration. At its core, U.S. medicine is composed of individual physicians who are paid each time they *treat* a patient for a *disease*, mostly on a fee-for-service basis. They may work in solo practice or in small or large groups, but their organizational framework differs little from those of preindustrial craftsmen: they are paid for piecework.

Sommer then proceeds to explore the very considerable consequences of such a reward system, for example, that they are paid to treat our disease du jour, not to keep us healthy. (I am indebted to a reviewer for drawing my attention to Sommer's book.)

So, what is the lesson here in our context? At the full system level, the critical issue is *resilience*, and Mukherjee (drawing on the work of David Simchi-Levi) suggests two critical metrics, *time to survive* (how long an enterprise can endure when there's a sudden shortage of some critical good) and *time to recover* (how much time it will take to restore adequate supplies of some critical good).

3.5. Failure to Think Through the Consequences of Using a Particular Metric

We might term this the Goldratt effect (Goldratt, 1990) and as noted earlier, Muller (2018) devoted a whole book to it. For example:

- It is still not uncommon for senior company executives to be rewarded on short-term performance, of which the Al Dunlap case at Sunbeam (2c) provides a classic instance. To counter such behaviors, the Stern Stewart approach to creating wealth based on Economic Value Added (EVA; see Ehrbar, 1999) contains a reward system. The system stipulates that CEOs' bonuses are dependent on company performance some years after they leave their posts, thereby forcing them to take the long-term interests of their companies into account.
- A survey of travelers on the suburban train system in Sydney identified 'on time arrival' as an important customer requirement. As a result, train drivers were given a performance target of '95% on-time arrivals at stations.' However, there was no further requirement that trains actually stop to allow passengers to board or alight from the train, with the obvious consequences.
- In an early version of so-called 'bulk-billing' at medical centers in Australia, patients were not charged fees. Doctors were paid according to the number of patients they saw, however briefly. This led to situations in which doctors saw up to 160 patients in a single day. The Australian government was forced to regulate that if a doctor saw more than 80 patients per day for 20 days, an explanation was required.
- High-rise residential developments on the outskirts of Cairo are distinguished by the fact that the top story is missing from many if not most of them. A city ordinance specifies that property taxes do not need to be paid until a building is complete.

3.6. Laziness: Measuring What's Easy Rather Than What's Potentially Useful

This practice brings to mind the old chestnut about the man searching for a lost wallet under the streetlight where he could see, rather than over in the dark area where he actually lost it. Many metrics that purport to be performance measures are, in fact, simply measures of activity. The earlier examples relating to the academic world (several bibliometrics, several entries in Table 2) fall into this category, but there are many others: number of meetings held, sales enquiries pursued, and so on.

3.7. Failure of Accountability, Responsibility and (Delegation of) Authority

This is a rather more complex issue. It hinges on the critical distinction between *accountability* and *responsibility* that is essential in developing a performance measurement system for an enterprise.

Suppose the board of an enterprise has just appointed you as chief executive officer. What does that mean? Well, it means you are in charge of everything—production, marketing, sales, recruitment, manufacture, delivery, resource management, budgeting, planning, and everything else. These constitute your areas of *accountability*. However, you do not actually do all of these things yourself. In reality, you delegate practically everything to your direct reports, *together with the authority to make decisions about whatever activities you delegated*. What's left, the things that won't get done if *you* don't come to work, constitute your actual areas of *responsibility*. (This distinction between accountability and responsibility is not well understood in English, let alone in French, Spanish, or Korean, in none of which languages a word for accountability even exists. And it did not exist in Japanese, until [Homer Sarasohn](#) introduced the word into the Japanese language in 1948, in the context of teaching Quality Management to Japanese business leaders; see Fisher, 2009.)

However, you need to report against your areas of accountability. Accountability cannot be delegated. And this is where a range of performance measurement problems occur. For example,

- The Exxon Valdez oil tanker struck Bligh Reef in the Prince William Sound, Alaska, on March 24, 1989, resulting in a massive oil spill that became one of the great man-made environmental disasters at sea. The captain was not on the bridge at the time. In the subsequent trial, he was convicted of a misdemeanor charge of negligent discharge of oil, but not held fully accountable for the disaster.
- A more subtle and insidious situation occurs when responsibility to carry out a task such as meeting a financial or market share target has been delegated to an individual, but without delegated authority to make decisions in relation to carrying out the task. The individual may then be held to account for missing targets despite not having full control over the means by which they might be achieved. In other words, they should not be held accountable for failure (or, for that matter, for success!).

3.8. Insistence on Objectivity and Repeatability

In this case, the contest is between metrics that always return the same value regardless of who performs the calculation, and metrics that are intrinsically a matter of judgment.

To anticipate the discussion of the next section, the ultimate performance measures are (in my view) *subjective*, not *objective*. We have already seen instances of this in relation to measuring the quality of academic research. Following the Tribus Paradigm, it is people's overall perception (that is, their subjective judgment) about whether they have received a good product or service that ultimately matters. For example, some investors argue that the only number one needs to know before investing in, say, a gold mining company, is EVA, the economic value added. EVA is the estimated economic profit in excess of the usual rate of return for an investment in this industry, and so is a hard accounting number. However, in the next section we argue that this is just one of a number of factors affecting an investor's overall perception of the real value. The lack of repeatability in the calculation of overall perceptual metrics is what frustrates accountants and others. It is easy

to count the number of published research articles, and the number does not vary from count to count, but not at all easy to assess people's overall perception of the quality of the published research, *all things considered*.

3.9. Using Rankings Not Ratings

This might appear to be unusual as a cause of problems in performance measurement. However, insistence on ranking people, institutions, investment funds, and so on, has created many problems. W. Edwards Deming famously highlighted these problems using the medium of his Red Bead experiment (e.g., Deming, 1982, or view Deming performing this experiment at https://www.youtube.com/watch?v=ckBfbvOXDvU) and Muller (2018) contains numerous examples.

General Electric and Motorola provide two striking examples from corporate life. As CEO of General Electric, Jack Welch instituted a policy of firing the bottom 10% of employees (whom he termed the C players) in each business unit every year (a variant of the old sailing ship practice of the last man to appear in response to the call 'All hands on deck' receiving the cat-o-nine-tails). And Motorola Corporation used to force senior officers to rank their people into quartiles each year. The fact is: in many situations, such as comparing ratings of the performance of several staff members over the last year, or the quality of universities on a range of criteria, or the performance of investment funds over 5 years, there may be no detectable statistical difference, even though the ratings are numerically distinct. Fisher (2021) explores this issue in detail in the context of how the world's universities are, or should be, compared.

Goldstein and Spiegelhalter (1996) discussed the use of so-called league tables when comparing institutional performance. A common intent for such tables is to provide a ranking. They queried whether the concept of 'value added' was really applicable in this context. Fisher (2019a) suggested how the concept might indeed be relevant.

4. One Approach to Making Progress

The goal of this section is to set out one systematic approach to problems of performance measurement, and to show how it can be applied in a number of practical situations. Some of the main elements have already been introduced. The key to this approach is to think carefully about the implications of the Tribus Paradigm introduced in Section 2, in the context of some examples. This section also serves as a very brief summary of material developed at length elsewhere; see, for example, Fisher (2013, 2019a, 2019b) for detailed explanations, technical and historical material, and credits. If no reference is provided in the discussion that follows, it can be assumed that relevant material is available from these sources.

4.2. Performance Measurement for an Enterprise

Being a director on the board of an enterprise brings with it numerous responsibilities known collectively as 'due diligence.' (Examples 2.2 and 2.3 represent classic failures in this regard.) Precisely what due diligence means varies greatly from jurisdiction to jurisdiction, but it invariably relates to safeguarding the interests of

the owners of the enterprise and, to varying extents, the welfare of its employees. For example, a director's due diligence might include how the enterprise conforms with legal and statutory requirements, and how it behaves toward its customers, the wider community; and so on.

Thus, a natural (statistical) question for a director to ask is: 'What information should I be seeing on, say, a monthly basis, that gives me confidence that I know how things are going, I know where we are heading, and I know where we need to focus attention next?' Which leads to the bigger question: What should be the scope and content of a (monthly) board report that will directors to be duly diligent in discharging their board responsibilities? In particular, board reports should facilitate asking the right questions of the executive leadership.

More generally, a Performance Measurement System for an enterprise should provide *everyone* in the enterprise with the quantitative information they need to do their jobs well. We start with the initial steps of the paradigm:

Step 1. What products or services are produced and for whom?

Step 2. How will 'quality,' or 'excellence' of the product or service be assessed and how can this be measured?

These steps constitute Stakeholder Analysis, which we encountered in Section 2 in the context of measuring the quality of academic research. Who are the stakeholders for, say, a factory manufacturing industrial chemicals, and what does 'quality,' or 'excellence' mean for them? It is convenient to categorize the stakeholders in distinct groups, as shown in Table 3, for reasons that will soon become apparent. The big challenge is, of course, to work out how to populate the third and fourth columns.

Table 3. Prototype stakeholder analysis for a chemical factory.

Stakeholder Group	Products or Services Provided	What Is 'Quality', Or 'Excellence'?	How Can This Be Measured?
Owners	Return on invested resources	?	?
Customers	Chemical products and advice	?	?
People	Jobs providing remuneration	?	?
Partners	Business via involvement in supply chain	?	?

| Community | Local employment, sponsorship for community activities, . . . | ? | ? |

Note. Each stakeholder group may itself need to be segmented. For example, there may be some very large customers as well as numerous small ones, with different product lines going to each. The third column is really important: it is the stakeholder who decides what Quality or Excellence means, and so it is the stakeholder who provides a guide to what should be in the fourth column. Note that the relative importance of these five groups is a matter of judgment for the board. However, each of the groups is making some sort of investment in the company, and expecting some sort of return.

Here, a key contribution was made by the leading strategic management thinker Richard Normann who, in the 1970s, introduced the concept that companies needed to "add value" for their customers, that is, to provide greater "value'' for customers than they could get elsewhere. Otherwise they would go elsewhere. However, Normann did not explain how "value,'' let alone "added value," might be measured. Some 10 years later, AT&T was forced to solve this problem, as noted in Section 2.1 and, in doing so, developed a very powerful and versatile improvement process (Kordupleski, 2003).

From meticulous study of their huge customer survey database (a triumph of data mining, before the term actually existed), Kordupleski and his team determined what really needed to be measured in order to obtain a reliable predictor of business outcomes. The critical quantity was Customer Value and in particular, Relative Customer Value, or Customer Value Added (CVA):

$$\mathrm{CVA} = \frac{Perceived\ worth\ of\ your\ offer}{Perceived\ worth\ of\ competitors'\ offers}$$

They interpreted 'value' to mean 'worth what paid for,' a customer's satisfaction with the *Quality of the product or service received*, balanced against their satisfaction with the *Price paid*. See Kordupleski (2003) for numerous case studies showing how well CVA performs as a predictor of superior business performance (for example, of market share, and return on invested capital).

So, one way to populate the third column of Table 3 is to devise a sensible concept for 'value' for each of the other stakeholder groups. Before doing this, we look at how we might actually measure value; this involves moving to the next step in the Tribus Paradigm.

Step 3. Which processes produce these products and services?

The AT&T approach to measuring value entailed elaborating its meaning in terms of its principal drivers and their attributes, which is readily understood by consulting the so-called Customer Value tree in Figure 1 relating to the customers of the chemical factory discussed in Section 4.1. (Quality practitioners will recognize this as a generalization of Quality Function Deployment.)
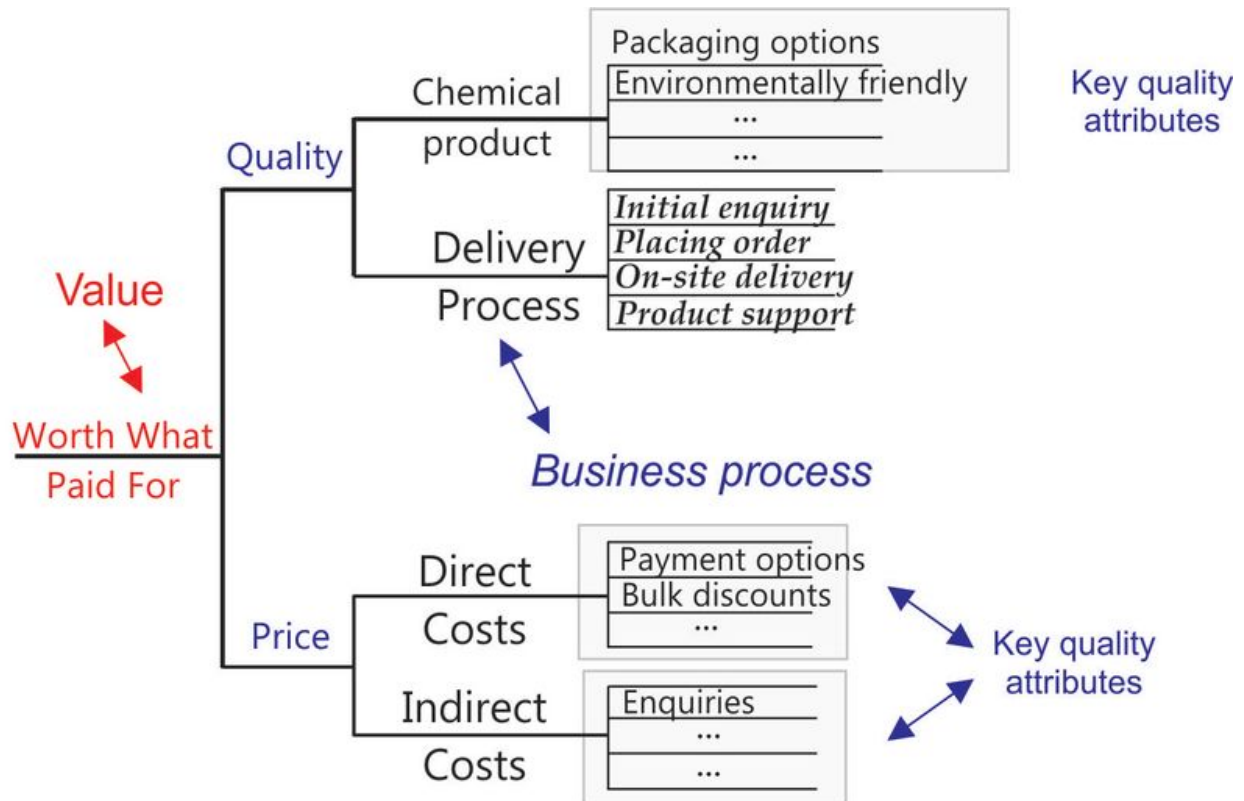
Figure 1. The Customer Value tree (in this example, a chemical factory) provides an elaboration of the customer's perception of overall Value in terms of its main drivers (their perception of the Quality of what they have received, balanced against their perception of the Price Paid) and their main attributes, which are determined from focus groups. The key to improving the outcome (Value) is to improve the business process from which it derives, and this process features prominently in the Value tree. This tree forms the basis for a Customer Value survey, data from which are critical input to an ongoing improvement process.

Figure 2 summarizes the whole improvement cycle (the Customer Value Management process is described in great detail in Kordupleski, 2003):

- use the tree to develop a survey instrument;
- conduct a market survey to capture respondent data;
- use the survey results to identify where to focus improvement efforts;
- make the improvements, communicate the improvements; and
- resurvey to confirm the efficacy of the improvement work and identify the next priorities.
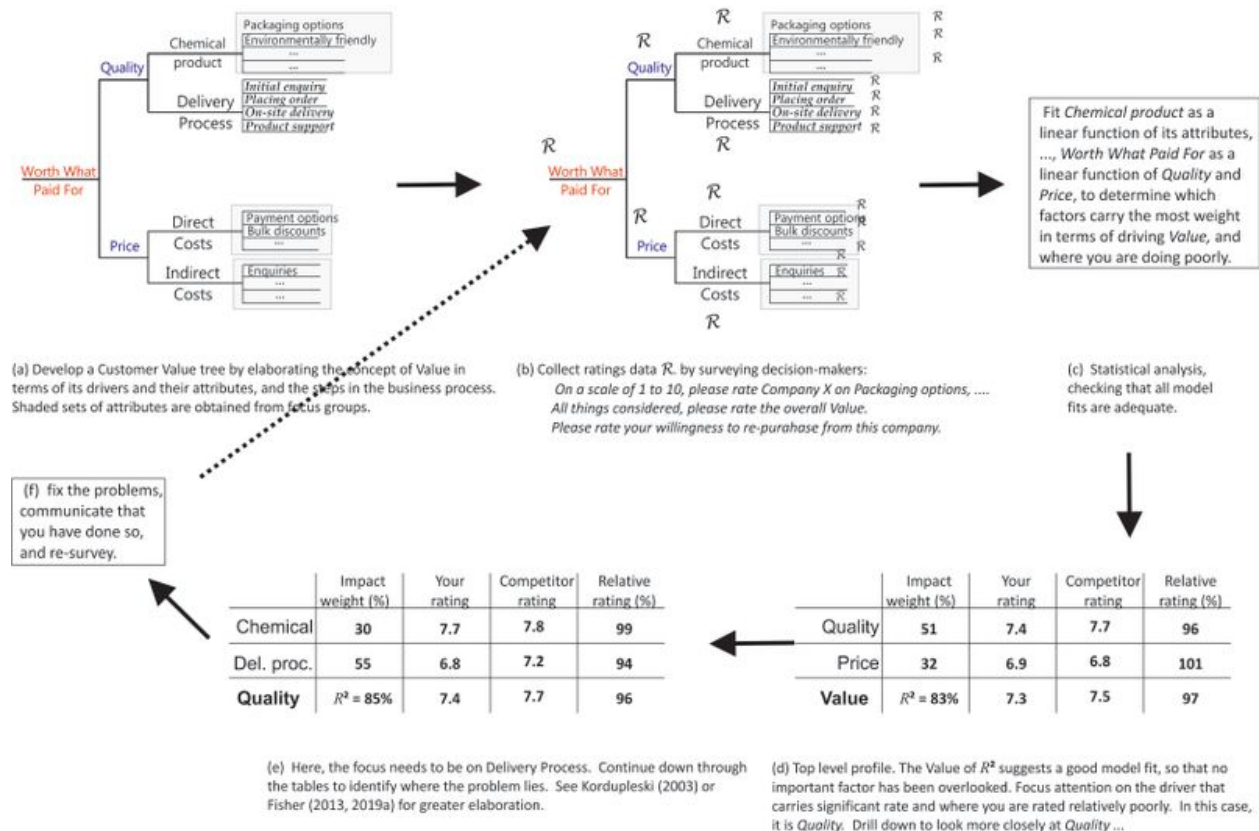
Figure 2. The Customer Value Management improvement cycle. Initially, advice from experts and market focus groups are used to develop a bespoke survey instrument based on the Customer Value tree. This provides a structured set of ratings that can be modeled hierarchically, yielding a set of tables that help identify the most important drivers of perceived value, and where specifically you are performing poorly. Having addressed the issues identified as priorities, they inform the market and resurvey. The data are synthetic, simply to illustrate the point. Numerous case studies can be found in Kordupleski (2003), for example.

There are some important points to note:

- The whole market is surveyed, if at all possible, so that *relative* performance can be assessed.
- The survey focuses on decision makers, that is, those responsible for making the purchasing decisions.
- Business impact questions need to be included at the end of the survey, to provide a connection between the overall value score and the business bottom line. For this purpose, respondents are also asked to rate their "Willingness to repurchase" (i.e., customer loyalty) and "Willingness to recommend" (i.e., customer advocacy). This information provides a form of internal benchmarking, although it is not as effective as external benchmarking. See Figure 3.
- The nature of the instrument provides confidence that no important factor affecting the overall perception of *value* has been omitted. Any such omission would be reflected in a poor model (as judged by an unacceptably low value of R^2 for models at one or more levels in the value tree).
- The resulting tables provide guidance about what needs to be fixed and in what order.
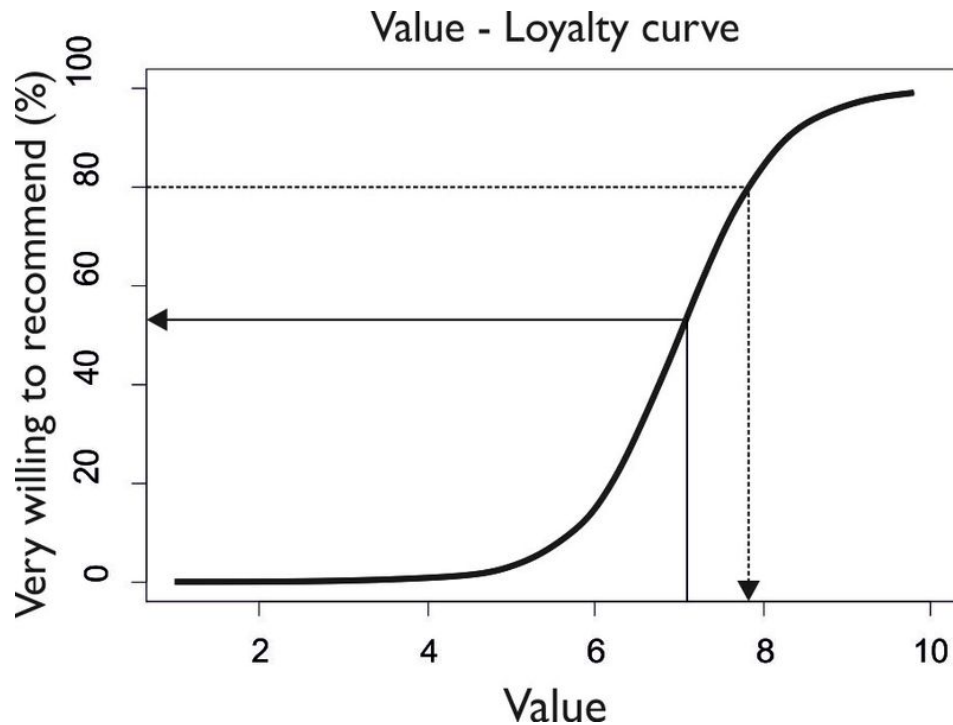
Figure 3. The Value–Loyalty curve connects the current value rating to a measure of loyalty. The current rating for value (7.3) corresponds to just 50% of customers being very willing to recommend you to others. To attain a loyalty of 80%, the value rating will need to improve to around 7.8.

Step 4. What has to be measured to forecast whether a satisfactory level of quality or excellence will be attained?

While the ultimate lead indicator is CVA, the only part under your control is how your enterprise is rated on value, quality, price, and so on. These, in turn, need internal hard (i.e., objective) metrics to act as lead indicators. See Kordupleski (2003) and Fisher (2013) for details of how to determine these in-process measures.

So now we are in a position to complete Table 3. Since the key aspects are captured in stakeholder value trees, this information is shown in Figure 4 for the other four stakeholder groups. (Note in particular that the representation for community value will depend very much on the issue at hand, and that the representation of partner value depends on the classification of the partnership as Strategic, Tactical or Operational.) The overall premise of this approach is that in the long run, an enterprise has to be seen to add value for all five groups, otherwise they will take their investments elsewhere and the enterprise will fail. It is easy to find examples of what happens when customers are treated badly, people are exploited, partners are deceived, or the community's trust that the company will 'do the right thing' is abused.
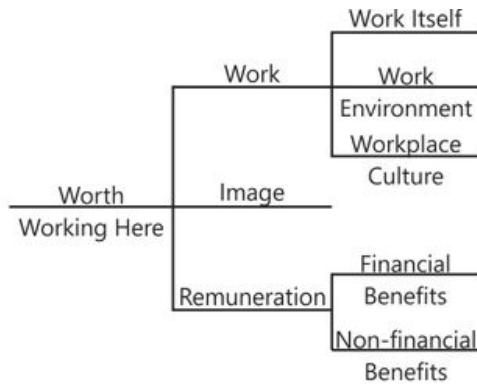
(a) People Value tree. Developed and tested successfully in a variety of academic, business and industry sectors 2001 – present.
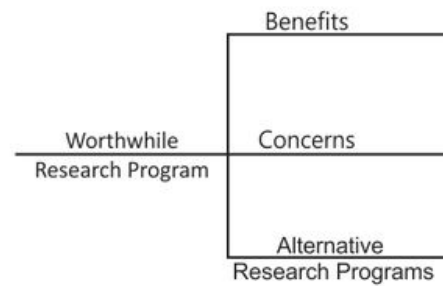
(b) One possible Community Value tree. Longitudinal studies during 2004 – 2011. with research agencies concerned with developing methods to manage invasive species. The concept of Value and its drivers is problem - dependent. See e.g. Fisher et al. (2008).
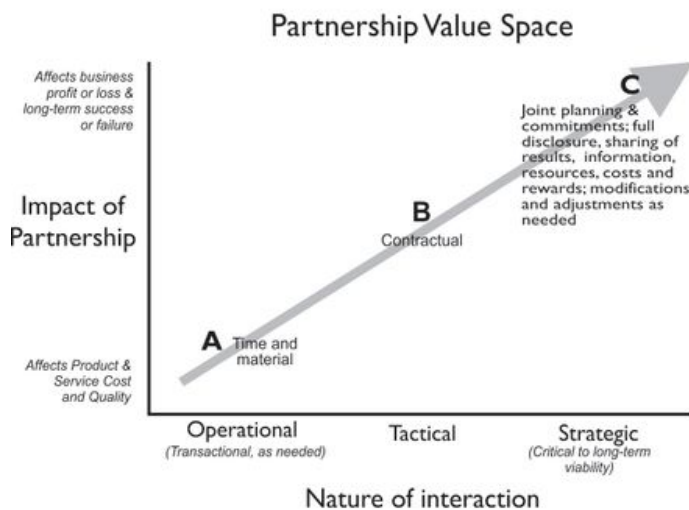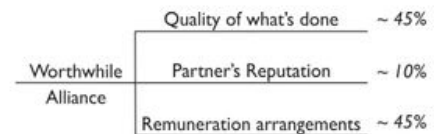
(c) Partnership Value trees. Studies with research collaborations that involved several agencies, during 2012 - 2017. Depending on the position of the relationship in Partnership Value Space, the Value tree will be very different. Relative weights associated with the branches are indicative. See Fisher & Peacock (2017).
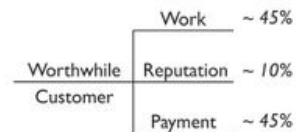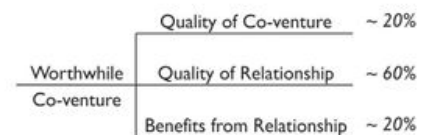
Figure 4. Representations of value for different stakeholder groups. For partners and for the community, it is context-dependent.

All of which enables us to summarize the Performance Measurement System. A simple analogy is the design of a car: you need to design the carriage-work, and then provide an engine. For a Performance Measurement System:

- the carriage-work is a *Performance Measurement Framework*, consisting of a set of Principles relating to alignment, process focus, and practicability; the Tribus Paradigm; and a Structure for performance measures as shown in Figure 5.
- the engine is a *Stakeholder Value Management* process, which is simply the analog of the process for managing Customer Value adapted to each stakeholder group.

Implementing such a system enables us to produce the sort of regular board and leadership reports for an enterprise that allow board members to act with due diligence·(see Fisher, 2013, 2019a, for details of such reports.)

Zones of Measurement
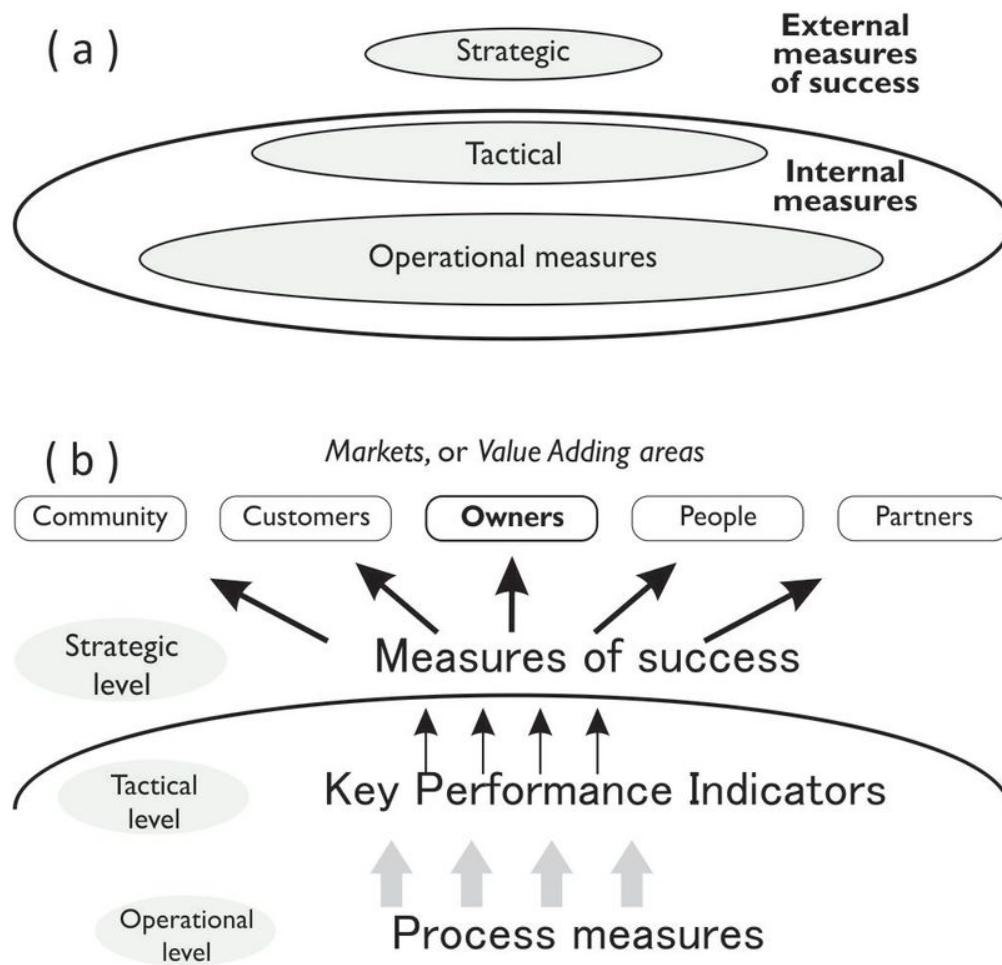
Figure 5. A structure for performance measures. (a) The three Zones of Measurement, providing structural alignment for performance measurements. (b) The five groups of stakeholders, for each of whom an enterprise needs to add value in order to survive and thrive in the long term. Measures of Success related to the value added for the stakeholder groups. They are so-called lag indicators, or measures associated with outcomes. The Tactical zone of performance measurement—lead indicators of success or Key Performance Indicators (KPIs) —the bottom line of the Tribus Paradigm. The Operational zone of performance measurement —operational measures relate to monitoring, controlling and improving processes that deliver, or support the delivery of, products and services to stakeholders.

4.2. Things to Bear in Mind in Tackling Other Problems of Performance Measurement

Evidently, there are elements of a generic approach to problems of performance measurement that emerge from what we've just described. In summary, these include

(a) thinking in terms of the Tribus Paradigm;

(b) proceeding from (a), the importance of stakeholder analysis;

(c) ensuring that there is a connection between the overall metric and the point where the ultimate impact needs to occur;

(d) where possible, embedding measurement in an ongoing process of continuous improvement; and

(e) viewing the measurement problem in an appropriate context of addressing a business issue.

4.3. Applying This Approach to Other Problems

Here, we mention two other classes of problems where this way of thinking has proved helpful. See Fisher (2019a) for some other performance measurement problems worth tackling.

Strategic Planning.

"Plans are useless, but planning is essential" is a sentiment generally attributed to Dwight Eisenhower and to Winston Churchill. Indeed, many, if not most, strategic plans fail (Fisher, 2018). And worse still, the time spent preparing them turns out to be wasted, notwithstanding. However, it is possible to design an approach that is somewhat robust against failure. A simple way to organize one's thinking about strategic planning is to start with a variant of the first step of the Tribus Paradigm:

Step 1. What products or services *will we need* to be providing in the coming years, and for whom?

The answers to this question allow you to frame strategic objectives for the various stakeholder groups so identified, and so devise associated strategies. Then, with this clear focus on stakeholders, the natural metrics for defining what it means to succeed with the objectives, and for monitoring progress toward achieving this, are simply those associated with the appropriate Stakeholder Value Management processes. This approach has been successfully deployed with a number of professional societies and in academic settings. See Fisher (2018) for details.

Managing the Culture of an Enterprise.

Culture matters.

After his experience as CEO at IBM, Louis Gerstner, Jr. commented (Gerstner, 2002, pp. 181–182):

> Until I came to IBM, I probably would have told you that culture was just one among several important elements in any organization's makeup and success—along with vision, strategy, marketing, financials, and the like...
>
> I came to see, in my time at IBM, that culture isn't just one aspect of the game—it is the game.
>
> In the end, an organization is nothing more than the collective capacity of its people to create value.

And recently, the [Australian Royal Commission into Financial Misconduct in the Financial Services Sector](#) commented at length about how poor culture was a root cause of the egregious behavior of Australia's largest banks toward their customers.

More specifically, safety culture matters.

The Chernobyl disaster in 1986 brought the issue into sharp relief (International Nuclear Safety Advisory Group, 1992):

> The accident can be said to have flowed from deficient safety culture, not only at the Chernobyl plant, but throughout the Soviet design, operating and regulatory organizations for nuclear power that existed at the time. Safety culture requires total dedication, which at nuclear power plants is primarily generated by the attitudes of managers of organizations involved in their development and operation.

Since then, almost every formal inquiry into serious safety incidents has concluded that organizational culture is a significant-to-major causal factor.

In the case of workplace safety, the traditional approach has tended to focus on collecting safety statistics on lost time injury frequency rates (LTIFR), damage to equipment, near misses, cost of claims, and so on. However, from any rational view, it is far more sensible—and vastly cheaper—to work 'upstream' on improving safety culture so that incidents will not occur in the first place because people are steeped in working safely. Again, this is something susceptible to monitoring and improvement using a variant of Stakeholder Value Management, by developing a Safety Culture tree structured similarly to the Stakeholder Value trees and implementing an associated improvement process. Such an approach is described in Fisher et al. (in press). And an important consequence of instituting such a process is that it provides people at all levels with the quantitative information they need to discharge their accountabilities and responsibilities in relation to safety of the workforce, and particularly the quantitative information needed at board level for directors to demonstrate due diligence in this regard.

5. Where to From Here?

5.1. Opportunities for Application of Performance Measurement

One very substantial sector largely untouched by general considerations of performance measurement is the public sector (Fisher, 2019a). Generally speaking, the only metrics used to monitor delivery of government programs relate to expenditure against budget, and project milestones being met. There are huge opportunities for improvement. Thus, an interesting question to ask of the government of any country, at the time of writing, is: What metrics are in place to give you confidence that the COVID-19 vaccination program will be carried out efficiently and effectively?

5.2. Opportunities to Develop Skills

Learning by doing can be a very effective way to get to grips with a new technology or a new way of thinking. In relation to this, a colleague, Professor John Bailer, has suggested it might be interesting to explore the application of performance measurement to the task of setting up a data science group (whether in a university or a company), or of establishing a data science consulting center.

In each case, the starting point is the same: a stakeholder analysis. Who are the different groups with a vested interest in such a group, and what are their needs? And so the process unfurls, with the end result being the sets of metrics needed to answer the questions: How are we going? Where are we heading? And where do we need to focus attention to improve?

(Being able to create a small expert system to codify this process is not beyond the realms of practical possibility.)

5.3. Research Possibilities

While potential applications of performance measurement abound, and are highly visible anywhere one looks, interesting technical issues worthy of exploration are not so easily identified. Typically, they emerge after wresting with important practical applications. What follows is, literally, a 'thought experiment' about a possible application of artificial intelligence (AI) algorithms.

Suppose that a large enterprise were to adopt the complete Performance Measurement System outlined in the previous section, and to set about capturing the full tree-structured perception data for their stakeholder groups. (This has been done in practice for individual stakeholder groups; see Fisher 2013, 2019a.) Full implementations of stakeholder management processes then lead to identification and collection of operational data for process monitoring and control (Kordupleski, 2003). A large enterprise capturing all these data streams in real time would rapidly accumulate a very substantial amount of multivariate temporal data.

One little-explored question in relation to stakeholder value trees is how the various attributes of the trees interact with other attributes in the same tree or with attributes in other trees, in a causal sense. As a trivial example, one people-value attribute for a call center operator might be *Having the skills and knowledge I need*

to do my job well. A customer-value attribute for someone contacting a call center might be *Problem solved on first call*. Thus, if the company works to upgrade the skills of the call center operator (with consequential improvement in the overall people value rating), there will be an improvement in the satisfaction of the customer and improvement in the overall customer value rating. AI algorithms might well produce rules that identify far more important causal connections that have very real impact on the business bottom line.

6. Concluding Remarks

Performance measurement is a fruitful area for research and development in need of smart people bringing original ideas. It is also important to appreciate that an understanding of context and purpose is essential in order to produce meaningful metrics. Using the approach described in this article entails application of the basic skills and understanding needed for process improvement, beyond the technical requirements for devising and computing metrics. A data scientist prepared to invest in developing a broader skill set has the opportunity to do work of significant impact.

Acknowledgments

Disclosure Statement

References

Adler, R., Ewing, J., & Taylor, P. (2009). Citation statistics. *Statistical Science*, *24*(1), 1–14. https://doi.org/10.1214/09-sts285

Arnauld, A. (1662). La logique ou L'art de penser: contenant outre les règles communes, plusieurs observations nouvelles propres à former le jugement. A Paris, Chez Charles Savreux, au pied de la Tour de Notre-Dame. Avec Privilege du Roy. (Logic or the Art of Thinking : Containing, besides the usual rules, several new observations useful in forming [one's] judgement. At Paris, at Charles Savreux's shop at the foot of Notre Dame. 1662. [Published] with the King's Privilege.)

Byrne, J. A. (1999). *Chainsaw: The notorious career of Al Dunlap in the era of profit-at-any-price.* HarperCollins.

Deming, W. E. (1982). *Out of the crisis*. Massachusetts Institute of Technology Center for Advanced Engineering Study. https://doi.org/10.7551/mitpress/11457.001.0001

Deming, W. E. (1994). *The new economics for industry, government, education* (2nd ed.). Massachusetts Institute of Technology Center for Advanced Engineering Study. https://doi.org/10.1002/qre.4680020421

Ehrbar, A. (1999). *EVA: The real key to creating wealth*. John Wiley & Sons.

Dransfield, S. B., Fisher, N. I., & Vogel, N. J. (1999). Using statistics and statistical thinking to improve organizational performance. *International Statistical Review, 67*(2), 99–150. https://doi.org/10.2307/1403389

Fisher, N. I. (2009). Homer Sarasohn and American involvement in the evolution of Quality Management in Japan, 1945–1950. *International Statistical Review, 77*(2), 276–299. https://doi:10.1111/j.1751-5823.2008.00065.x

Fisher, N. I. (2013). *Analytics for leaders: A performance measurement system for business success*. Cambridge University Press. https://doi.org/10.1017/cbo9781107053779

Fisher, N. I. (2018). Stakeholder value as an organising principle for strategic planning, with application to a university department. *Journal of Creating Value, 4*(1), 168–177. https://doi.org/10.1177/2394964318771251

Fisher, N. I. (2019a). A comprehensive approach to problems of performance measurement. *Journal of the Royal Statistical Society: Series A, 182*(3), 755–803. https://doi.org/10.1111/rssa.12424

Fisher, N. I. (2019b). *Walking with giants: A research odyssey* [Video]. 2019 Deming Lecture, Joint Statistical Meetings, Denver, Colorado, July 31, 2019. https://ww2.amstat.org/meetings/jsm/2019/webcasts/index.cfm; Presentation material: https://www.amstat.org/ASA/Your-Career/Awards/Deming-Lecturer-Award.aspx.

Fisher, N. I. (2021). *Assessing the quality of universities: A Gedankenexperiment* [Manuscript submitted for publication]. School of Mathematics & Statistics, University of Sydney.

Fisher, N. I., & Kordupleski, R. E. (2019). Good and bad market research: What's wrong with net promoter score, and why. *Applied Stochastic Models in Business and Industry, 35*(1), 138–151. https://doi.org/10.1002/asmb.2417

Fisher, N. I., Lunn, P., & Sasse, S. M. (in press). Enhancing value by continuously improving enterprise culture. *Journal of Creating Value, 7*(2), 232–254. https://doi.org/10.1177/23949643211044048

Fisher, N. I., & Vogel, N. J. (2017, October 1). Obituary: Myron Tribus. *AMSTATNEWS*. http://magazine.amstat.org/blog/2017/10/01/obituary-myron-tribus/

Gerstner, L. V. Jr. (2002). *Who says elephants can't dance?* HarperCollins.

Goldratt, E. M. (1990). *The haystack syndrome: Sifting information out of the data ocean*. North River Press.

Goldstein, H., & Spiegelhalter, D. J. (1996). League tables and their limitations: Statistical issues in comparisons of institutional performance. *Journal of the Royal Statistical Society: Series A, 159*(3), 385–443. https://doi.org/10.2307/2983325

International Nuclear Safety Advisory Group. (1992). *INSAG-7, The Chernobyl accident, updating of INSAG-1: A report by the International Nuclear Safety Advisory Group*. https://doi.org/10.1016/0160-4120(93)90296-t

Kordupleski, R. (2003). *Mastering customer value management*. Pinnaflex Educational Resources.

Mukherjee, S. (2020, May 4). What the coronavirus crisis reveals about American medicine. *The New Yorker*. https://www.newyorker.com/magazine/2020/05/04/what-the-coronavirus-crisis-reveals-about-american-medicine

Muller, J. Z. (2018). *The tyranny of metrics*. Princeton University Press. https://doi.org/10.1111/gove.12510

Price, F. (1984). *Right first time*. Wildwood House Limited. https://doi.org/10.4324/9781315244143

Reichheld, F. F. (2003, December). The one number you need to grow. *Harvard Business Review, 81*(12), 46–54. https://hbr.org/2003/12/the-one-number-you-need-to-grow

Sommer, A. (2009). *Getting what we deserve: Health and medical care in America*. The Johns Hopkins University Press. https://doi.org/10.1093/aje/kwp405

Zingales, L. (2020, January 9). Friedman's Principle, 50 years later. *Promarket.* https://promarket.org/2020/09/01/friedmans-principle-50-years-later/

Footnotes

1. The most senior officers in a firm, often with titles including the word 'Chief.' ↩