ARC Engagement and Impact Consultation Paper: Response

June 2016

The Cooperative Research Centres Association represents all Australian Cooperative Research Centres (CRCs). In addition, the Association has universities, companies and research groups as Affiliate and Associate Members.

ARC Engagement and Impact Consultation Paper: Response
Febuary 2016

For futher inquires contact:

Dr Nick Fisher
Principal
ValueMetrics Australia
PO Box 1049, North Sydney NSW 2059
02 9922 1623
http://www.valuemetrics.com.au

Dr. Tony Peacock
Chief Executive Officer
CRC Association
1/10 Bourke Street BARTON ACT 2600
02 6273 0624
www.crca.asn.au

This work: ARC Engagement and Impact Consultation Paper: Response, by N I Fisher & A J Peacock, is free of known copyright restrictions.

ARC Engagement and Impact Consultation Paper: Response

N I Fisher[1] & A J Peacock[2]

Summary

The Cooperative Research Centres Association is extremely supportive of an Engagement and Impact Measure being adopted to stand beside the Excellence in Research Australia (ERA) measure. The quality of Australian research is obviously of critical importance, but we believe the current ERA system tends to push universities and individual researchers to an over emphasis on international journal publication and an under emphasis on the use of their findings to make a difference in the world. Our support of an Engagement and Impact Measure is aimed at balancing the incentives to researchers to ensure they make the necessary efforts to maximise the impact of their work.

The quality and type of engagement is obviously of critical importance and therefore we are providing the following paper in an attempt to ensure methods of evaluating engagement and impact are fully explored. The CRC Association wants a simple Engagement and Impact Measure put in place quickly, but we are very conscious that too much speed or too much simplicity could result in the wrong incentives being given to researchers.

This paper explores a simple method for assessing Engagement and Impact that has been successfully employed for research groups in Australia. Our method is more complex and costly than a simple index of currently collected data, but less so than the gathering and assessment of case studies. If deemed worthy of adopting on a wide spread basis, the method could be scaled-up with an initial investment but then it would be relatively resource light to capture stakeholder perception of Engagement and Impact on an ongoing basis.

1. Introduction

Performance measurement arises in a wide variety of contexts. Consider, for example, the following areas:

1. Performance measurement for an enterprise – what sorts of metrics are needed at each level in the enterprise from Board to shop floor to enable people to do their jobs well, and how can these metrics be obtained and used to support good decision-making?
2. Measuring workplace culture
3. Measuring research quality / engagement / impact
4. Managing efficient and effective implementation of public policy
5. Ranking universities / Business Schools / …

[1] Principal, ValueMetrics Australia, and Visiting Professor of Statistics, University of Sydney. nif@valuemetrics.com.au

[2] Chief Executive Officer, Cooperative Research Centres Association. Tony Peacock. tpeacock@crca.asn.au

There is still no widely-accepted methodology providing a systematic way of implementing a performance measurement system to address any of these problems. The term 'methodology' is taken to mean a set of principles, structures and processes of the sort being sought by the ARC for assessing engagement and impact.

However, there has been progress in developing such methodology to address the first of these contexts (Fisher 2013). Further, key aspects of the methodology have been adapted successfully (as judged by case studies) to the contexts of Workplace Culture, and the Quality of Applied Research groups, suggesting that generic approach developed in the enterprise measurement context is worthy of serious consideration in the current context. A very brief overview of the methodology is provided in the Appendix, and the proposals will be explicated in terms of the concepts and line of argument introduced there.

To exemplify the way the methodology might be adapted, the discussion will focus on some specific examples of individual and group research activities, and the people or groups for whom research results are intended to provide direct benefits. Here are some diverse examples of researchers and research groups:

1. A linguistics expert studying grammatical structures in Ancient Arabic.
2. A pure mathematician attempting to prove the Riemann hypothesis.
3. A group seeking to develop a practical means of generating energy by nuclear fusion.
4. A research centre seeking to develop improved methods of managing invasive animals, such as European carp in Australian waterways
5. A research group seeking to develop the legal basis for the rights of refugees arriving in Australia by boat.

Section 2 of the submission considers the issue of defining and measuring "Impact" and Section 3 looks at the corresponding issue for "Engagement". Section 4 contains some final comments.

2. Defining "Impact" and "Engagement"

Defining "Impact"

One clear assumption of the material presented in the Appendix is that **Impact** is a matter of perception … of *perceived Value*. It is not a hard number such as a profit of $2.7M, or establishing an original research result. Those are *outputs*, whereas **Impact** relates to *outcomes*. For example,

- The linguistics expert in Ancient Arabic (1) publishes research for the benefit of others working in the same field and also in cognate fields for other languages. The notion of Value for, say, a linguistics expert specialising in Russian would include aspect such as: *Published in a language that I read*, and *Clearly explained*. So, unless the expert in Ancient Arabic is using a reasonably international language (English, Spanish, …) to report results, part of the target audience is disenfranchised.
- There will be many different stakeholders interested in the success of the group working on nuclear fusion – the university where the research takes place, venture

capitalists providing funding, the research community, the general community in terms of a cheap long-term solution to their energy needs, … . The list is very long, so for the moment, we just focus on the venture capitalists. Their overall Investment Value tree would have a range of Attributes apart from the technology being developed, not least, successful applications for patents. The **Impact**, or Value, will be based on their perception of how well all of these aspects have been handled.

- There will also be many different stakeholder groups interested in the work of the research centre seeking to develop improved methods of managing invasive animals. Take, for example, the Invasive Animals CRC research into managing carp using a herpes virus. The consequences of the research program may or may not be beneficial!
    - The general community benefits from cleaner waterways and the improved conditions for native species using the water.
    - On the other hand, people who make a living from overseas commercial sales of carp products will have their business seriously affected.

This then provides a way to approach the question of defining **Impact**. There are two strands to the definition:

1. Identifying each group for whom **Impact** needs to be defined.
2. For that group, defining what *Value* means for the group. In some cases, this may be described (as in Customer Value) by *Worth What Paid For*. For the linguistics expert specialising in Russian, it may be *Useable research results*. For the venture capitalist it may be *Worthwhile investment*. And so on. Value will be stakeholder-dependent and in each case will have a number of drivers and attributes, so that the overall perception of Value is the result of the collective influenced of many factors.

Defining "Engagement"

The same issues occur as for **Impact**: who's asking, and what is their concept of (for want of a better term) **Worthwhile engagement**.

The ATSE formulae (ATSE 2016) measure engagement in purely monetary terms – total external revenue – normalised in various ways. (We note that these quantities can be measured down to FoR code level.) It may be the case that, for one or two stakeholder groups, the notion of **Engagement** is entirely (or at least, essentially) captured by one or two simple accounting numbers. However, this is a significant assumption that needs to be tested. For example if, as a consequences of the **Engagement**, a major tech investor decides that they will never work with a university again, shouldn't that also be captured in the measurement of **Engagement**?

Thus, our argument is that the ultimate measures of **Impact** and **Engagement** are both perceptual and stakeholder-dependent. However, that does not mean that there are no practical ways to make the measurements.

3. Measuring "Impact" and Engagement"

Here, we present our ideas for the case of **Impact**, but they are readily adapted to studying **Engagement.**

Defining the measures

It may be helpful to use the methodology outlined in Section A2 of the Appendix to be explicit about the characteristics of the metrics being sought, as a basis for making a judgment about competing proposals. For example, Figure 1 is illustrative of what be regarded as desirable.

Evidently, the ATSE measures of engagement place significant weight on Price. However, the ATSE measures have defined engagement purely in financial terms. Thus, they provide no insight into what might be the critical influences on the overall metric, which could provide a basis for improvement activities.

The alternative to purely objective metrics is to use measures of perception that involve some form of peer assessment. Of course, the peer assessment may use, as inputs to the decision-making process, a range of objective metrics such as the ATSE metrics, bibliometric and other data, but the final decision is a matter of judgment by experts. This is the general approach taken by the REF scheme in the United Kingdom. Whilst a peer-assessment-centred method may be more expensive to adopt, it may result in better decision-making, because of its superiority as judged by the criteria in Figure 1. A famous quote by John W Tukey, one of the greatest statistical scientists of all time, is apposite:

> "Far better an approximate answer to the right question, which is often vague, than an exact answer to the wrong question, which can always be made precise." (Tukey 1962, page 13.)

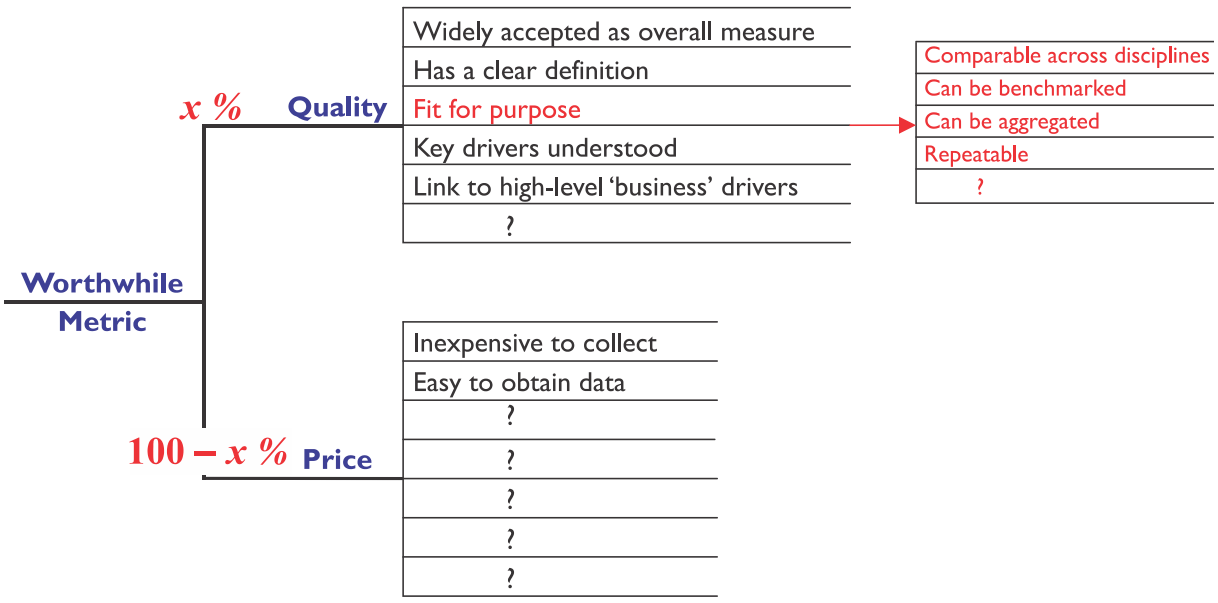

Figure 1 Tree representation of the characteristics being sought in measures for Impact and Engagement. Some aspects of *Fit for purpose* have been elaborated. If the key drivers are understood, this provides the basis for identifying where improvement efforts need to be focused.

Here, we provide two examples of how the methodology of Section A2 might be applied to measure the **Impact** of research, from the perspectives of two very different categories of stakeholder. We then make some brief general comments on a process to implement collection of the measures, informed by a recent case study at an Australian university.

Example 1. Suppose some venture capitalists are funding a university group carrying out research into nuclear fusion. Whilst not the only important stakeholder group, the investors clearly comprise an important one, and they will have a unique perspective about the **Impact** of the research. The investors' judgment about whether or not this project has been a worthwhile investment may plausibly be represented as shown in Figure 2.

This model can be used both to monitor investment perceptions during the conduct of the research (*e.g.* at a significant project milestone), and after completion of the project to obtain an overall metric[3] for **Impact**.

Further (*cf.* Appendix A2), the investors can be asked to provide ratings for high-level business drivers such as

BD1: Willingness to recommend this research group to others
BD2: Willingness to re-invest with this research group on other projects

In fact, this leads to a method of calibrating the **Impact** (Worthwhile investment) scores, apart from their position on the scale of 1 (Poor) to 10 (excellent). For example, by accumulating data on **Impact** scores and BD1 scores from a number of projects, it is possible to create a curve such as shown in Figure 3.

Note that the higher-level branches of the **Impact** tree in Figure 2 are generic across the Investor community, and so suited to benchmarking, and one would expect significant commonality for the lower-level factors as well. In fact, consultation with a number of investors may well lead to the conclusion that a generic **Impact** tree can be developed.

There will, of course, be other stakeholders for whom the research group hopes to make significant impact (externally, the wider research community and, longer-term, the wider community through cheaper and cleaner energy production; internally, parts of the university itself). Corresponding **Impact** trees can be developed for these stakeholder groups, with corresponding Impact – Loyalty curves and commonality of higher-level branches across research groups.

[3] Following the method of Appendix A2, each investor would be asked to provide ratings for all branches of the tree, as well as percentages indicating relative importance of the drivers at each level. Then the metric for **Impact** is simply the average rating assigned to "Worthwhile Investment".

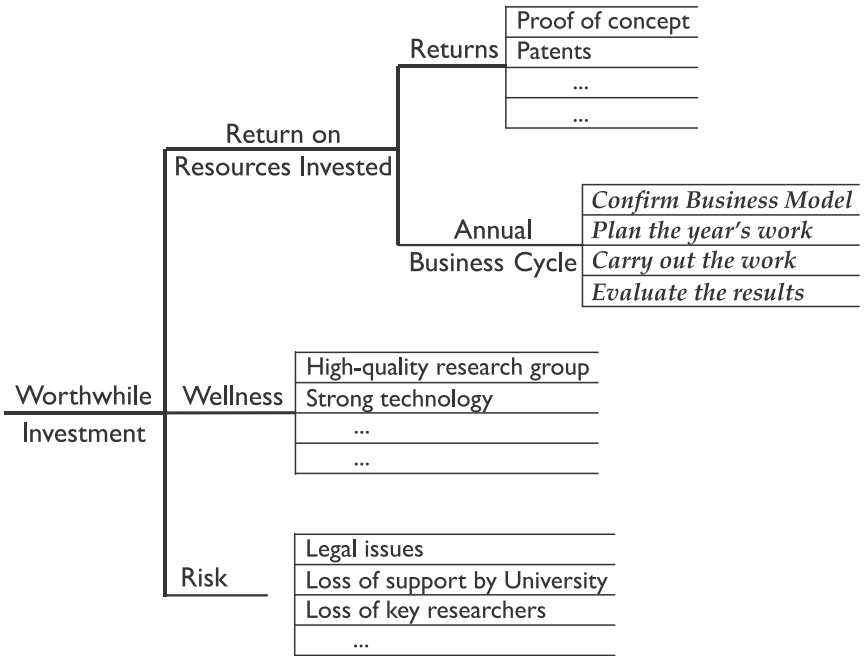

Figure 2 Possible model for Impact on Investors, for a small number of investors in a university research group. The higher-level branches are generic across the Investor community, and so suited to benchmarking, and one would expect significant commonality for the lower-level factors as well.

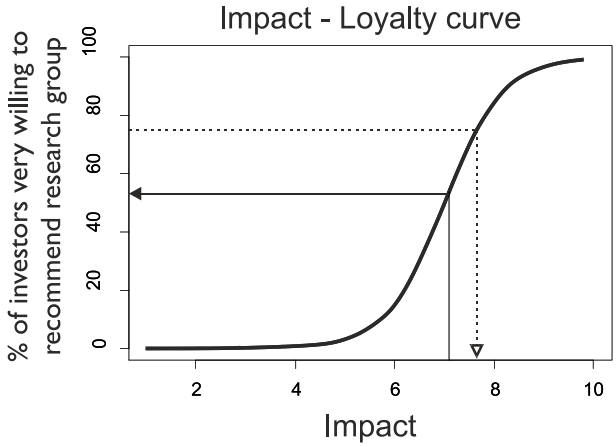

Figure 3 An Impact–Loyalty graph shows the relationship between Impact, on a 10-point scale, and a measure of Loyalty, in this case, the percentage of people who are very willing to recommend a research group to others. In this example, the current Impact rating of 7.1 corresponds (solid arrow) to about 63% of investors being very willing to recommend groups of this calibre to others. If the 12-month target for Loyalty is chosen as 75%, this means that the Impact score has to be increased to about 7.65.

Example 2. Here, we consider a research centre seeking to develop viral methods of managing invasive animals, such as European carp in Australian waterways. Long-term beneficiaries would include the Australian community because of the prospective environmental benefits of eradicating or dramatically reducing carp numbers (much cleaner water, more abundant native species such as Murray Cod, …), apart from improved fishing for recreational fishermen, *etc*. Now, the realisation of benefits may not occur for decades, although they are predictable by environmentalists. How should the (potential) impact on the community, possibly the most significant stakeholder, be assessed?

A different sort of **Impact** tree needs to be developed, possibly structured something like that shown in Figure 4 (but this would be subject to consultation with the community):

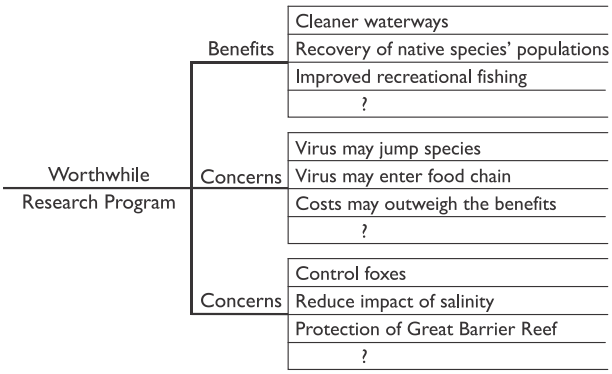

Figure 4 Possible model for Impact on the community, for a research project with the benefits being realised some time in the future. The higher-level branches are generic across such projects, and so suited to benchmarking.

Other comments under Example 1, including analogues of Loyalty and of the Impact – loyalty curve (Figure 3) can then be developed.

A process for putting the measures in place

A process somewhat comparable to the UK REF approach has been successfully trialled in one Australian university to assess and calibrate the quality of applied research groups whose work is not necessarily treated satisfactorily by ERA (for reasons canvassed in the consultation paper). The process features:

- a small Peer review team
- a brief structured submission by the group being assessed
- impact assessments by major stakeholders along the lines described in Examples 1 & 2
- ready identification of where to focus attention to make improvements in Impact and Engagements
- very low impact on the resources of people at the university or on the review team.

More detailed information can be supplied if required.

4. Discussion

Comments on the guiding principles

a. **Robust and objective** —*objective measures that meet a defined methodology that will reliably produce the same result, regardless of when and by whom the principles are applied*.

Comment: In light of the arguments provided earlier in this submission, the term "Objective" is not appropriate. The highest-level measures of Impact and Engagement are almost certainly based on expert judgment. That need NOT prevent them from being essentially reproducible.

"Regardless of when and by whom the principles are applied" is also not appropriate. If some form of peer assessment is used, then the assessors must be recognised as being of good standing in the area.

b. **Internationally recognised**—*while not all indicators will allow for direct international comparability, the indicators must be internationally recognised measures of research engagement and impact. Indicators must be sensitive to a range of research types, including research relevant to different audiences (e.g. practitioner focused, internationally relevant, nationally- and regionally-focused research).*

   Comment: If no-one is doing it well (which may well be the case), then this criterion is questionable unless it is interpreted as "acceptable to the international community".

c. [**"Time bound"** should read Time-bounded]

Investment needed

The approach advocated in this submission requires the investment of significant resources at the beginning of the process. This is often the case: the quality of the outputs and outcomes are critically dependent on the quality of the initial specification and design work. That said, with the various **Impact** and **Engagement** trees in place (and available as Web-based instruments), it is a simple and resource-light matter to capture stakeholder perception data on an ongoing basis.

Appendix: Performance Measurement for an enterprise – brief overview

Fisher (2013) suggested that a performance measurement system for an enterprise should provide:

- a concise and comprehensive overview of the health of the enterprise;
- a quantitative basis for selecting improvement priorities; and
- alignment of the efforts of the people with the mission of the enterprise.

The system is based on a **Performance Measurement Framework** (PMF®) for performance measurement that specifies the sorts of performance measurements needed, together with a general enabling process, **Stakeholder Value Management**, that shows how to put the performance measurements in place and use them to best effect.

A1. The Performance Measurement Framework

This comprises a set of *Principles*, a *Paradigm* for performance measures, and a *Structure* for performance measures.

The Principles underpinning the system are:

- Alignment – the enterprise's approach to measurement encourages alignment of people and systems with the organization's mission, vision and goals.
- Process and systems thinking – measures should be linked appropriately with system and process monitoring, control and improvement
- Practicability – at any level in the enterprise, there is a straightforward procedure for identifying the sorts of measures that need to be collected, and what needs to be reported.

Then there is a basic thought process for developing measures in any given situation, known as the *Tribus Paradigm*:

1. What products or services are produced and for whom?
2. How will 'quality', or 'excellence' of the product or service be assessed and how can this be measured?
3. Which processes produce these products and services?
4. What has to be measured to forecast whether a satisfactory level of quality will be attained?

Finally, and of particular relevance to the Consultation paper, there is a Structure for performance measurement. The whole system takes a *stakeholder view of life*: the enterprise will only survive and thrive in the long term if it *creates and adds value* for its key stakeholders, in other words, for those people or groups who make some form of investment in the enterprise, and so expect some sort of return. These key stakeholders can be classified as

- *Owners* (*e.g.* shareholders of a publicly-listed company; the government, for some universities; a venture capitalist, for some research groups)
- *Customers* – those for whom work (*e.g.* teaching, research, …) is done

- *People* – those who work for the enterprise (*e.g.* faculty, teaching assistants, administrators … for a university)
- *Partners* – people or groups with whom the enterprise collaborates (*e.g.* research partners), possibly including strategically important suppliers (*e.g.* for a university, undergraduate or graduate programs elsewhere)
- *Community* – which could be the community of research groups, or the general public, whose support for, say, research into genetically modified viruses to control invasive species may ultimately determine whether the results of the research are permitted to be deployed.

Thus, *the starting point of this approach to performance measurement is what it means to be successful, for each of these groups*. In the language of this system, success equates to *creating and adding superior value* compared with what each group might obtain elsewhere. The meaning of the term "Value" is explored below.

Of course, the groups are not necessarily of equal importance. Also, and very importantly, *each stakeholder group may comprise several sub-groups which have widely differing views about value*!

A summary of the structure for performance measures is displayed in Figure A1. Figure A1(a) shows three basic *Zones of Measurement* – Strategic, Tactical and Operational – and Figure A1(b) elaborates the nature of each zone.

- Strategic zone contains measures of Success, and corresponds to where the ultimate impact of the enterprise is judged by the stakeholders. However, these measures are *outcome measures* – the consequences of past actions by the enterprise – and so they are necessarily *lag indicators*.
- The Tactical zone is critical for leadership, as it contains measures whose purpose is predict where the enterprise is heading: *is it on target to be successful with its stakeholders*? In other words, it comprises *lead indicators of success*.
- The Operational zone comprises all the lower-level performance measures required for monitoring, controlling and improving products and services provided to Customers and other stakeholders.

According to the system, measurements are needed in all zones. The starting point has to be the Strategic zone (*cf.* the Tribus paradigm) to ensure alignment of the enterprise with its goal of being successful with all stakeholders (*cf.* the Alignment Principle) by specifying the target (for each Stakeholder group). The next question is then: how to obtain and use the measurements? For this, a generic process is needed.

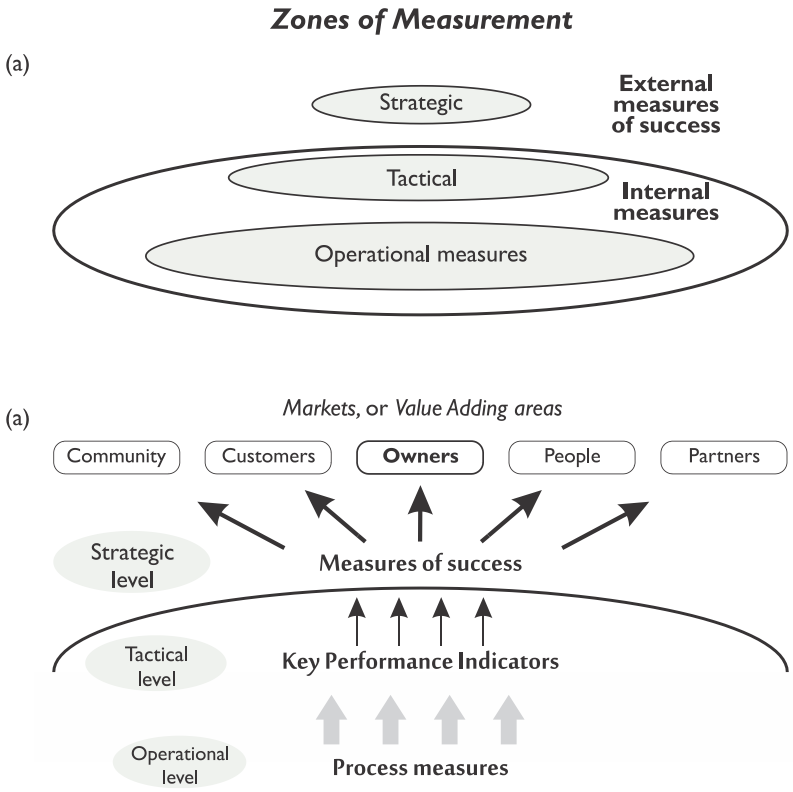

Figure A1 The two figures show the Performance measurement Framework, with its focus on five stakeholder groups, and with three distinct Zones of Measurement delineated. In the Strategic Zone, Measures of Success are outcome, or lag measures, a consequence of the past efforts of the enterprise. The Tactical Zone is of more interest to the leadership of the enterprise in terms of management, because the Key Performance Indicators are lead indicators of future success. The Operational Zone comprises the usual metrics for monitoring, controlling and improving processes and so improving KPIs.

A2. The Stakeholder Value Management process

As described by Fisher (2013), this is based on a process for managing Customer Value developed at AT&T in the mid-1980s in response to a business crisis (see also Kordupleski 2003 for a discussion of its beginnings). It has the following benefits: Customer Value Management (CVM) is a proven process used by leading organisations world-wide that provides, *inter alia*, lead indicators of business results and actionable Board and senior executive reports, and helps identify priorities likely to have the biggest impact on the business bottom line. We provide a simple example from a different context purely for illustrative purposes.

The starting point is to define the notion of *Value* for Customers of a product or service. This is taken to mean *Worth What Paid For*, the trade-off between the customer's satisfaction with the *Quality* of the product or service received balanced against satisfaction with *Price* paid. In many cases, satisfaction with *Image* becomes a third so-called driver of *Value*. Each of the drivers can then be further elaborated, as shown in Figure A2(a), where *Quality* has its own drivers the particular *Product* or *Service* being offered, and the *Delivery Process*, or sequence of customer experiences during provision of the product or service, and likewise Price can be decomposed into *Direct Costs* and *Cost of Doing Business*. Attributes of the drivers are determined from market focus groups.

Data are obtained by asking respondents to rate the performance of their supplier on a 10-point scale, where 1 is *Poor* and 10 is *Excellent*, starting with the attributes of the *Product* or *Service*, followed by an overall rating of *Product/Service*, and so on, until eventually an overall rating for *Value* is assigned. This leads to a structured set of ratings (Figure A2(b)) that can then be modelled and analysed in hierarchical fashion, by fitting *Product/Service* as a function of its attributes, *Delivery Process* as a function of its sub-processes, *Quality* as a function of *Product/Service* and *Delivery Process,* ..., and finally, *Value* as a function of *Quality*, *Image* and *Price*.

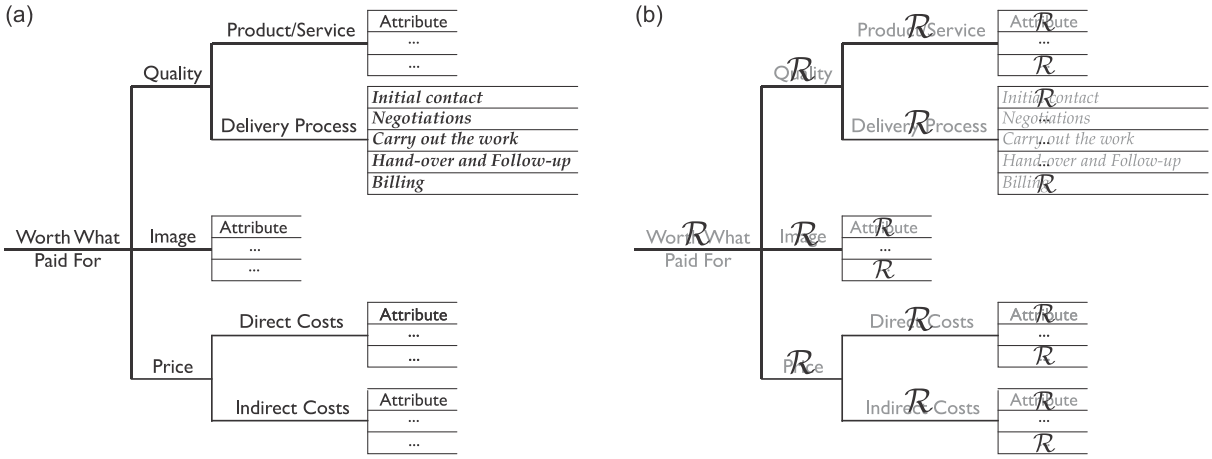

Figure A2 (a) Proto-typical structure of a Customer Value tree. In this representation, overall *Value* (described as *Worth What Paid For*) has three main 'drivers': *Quality*, *Image* and *Price*. *Quality* is represented as the *Product or Service Received*, and the *Delivery Process* comprising the sequence of Customer experiences in receiving this *Product or Service*. *Price* has as its Drivers both *Direct Costs* and *Indirect Costs* (or *Cost of Doing Business*). The small sets of 'attributes' associated with each of the main branches are determined from market focus groups. (b) The survey produces ratings (*R*) on a 10-point scale, with a hierarchical structure Figure 3(b)).

These data can now be modelled and analysed statistically to ascertain which branches of the *Value tree* carry the most weight in terms of influencing overall *Value* and are, at the same time, rated poorly, which provides the basis for focusing improvement activity. (An important side benefit of this approach to satisfaction surveys is that it is also possible to check, statistically, that no important factor has been omitted from the survey.) The results can then be reported as shown in Table A1 for the top level of the tree.

	Impact weight (%)	**Mean Rating** (s.e. ± 0.05)		**Relative rating** (%)
		This company	Competition	
Quality (Q)	32	7.7	7.6	101
Image (I)	15	7.1	7.3	97
Price (P)	40	6.8	7.2	98
Value (V)		7.1	7.5	97

Table A1 A top-level Customer Value competitive profile. The overall Value score is 7.3, compared with the competition which averages 7.5, so for this enterprise, the Relative Value, or Customer Value Added (CVA) score is below par at 97%. (Par would be a CVA score in the range 98 – 102, allowing for the variability in the data.) The ratings are on a 10-point scale, where 1 = Poor and 10 = Excellent. Based on these results, attention should be focused initially on Price and its drivers.

Of course, improving *Value* is of no interest unless *Value* itself can be connected to some higher-level business driver. This can be accomplished by including appropriate questions in the survey, relating to, say, *Willingness to repurchase*, or *Willingness to recommend the product or service to others*. This allows for a graph like Figure A3 to be constructed.

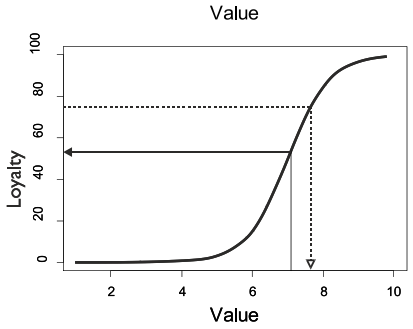

Figure A3 A Value–Loyalty graph shows the typical relationship between Value, on a 10-point scale, and the percentage of people who are very willing to show some specific form of Loyalty, such as being very willing to recommend an enterprise to others, or very willing to re-purchase. In this example, the specific score Value = 7.1 corresponds (solid arrow) to about 63% of people being very willing to recommend (*i.e.* providing a rating of 8, 9 or 10 on a 10-point scale). If the 12-month target for Loyalty is chosen as 75%, this means that the Value score has to be increased to about 7.65.

This technique can then form the basis for an ongoing process of continuing improvement of Customer Value:

1. Understand what Value means to a customer, and identify the key drivers of Value, and how to connect Value to higher level business drivers.
2. Put measures and measurement processes in place to acquire data.
3. Acquire and analyse Value survey data.
4. Identify the improvement priorities that will have the greatest impact on Value, and hence on your business.
5. Make the improvements and communicate the improvements to the customer.
6. Go to (3).

This same line of reasoning can be applied to define the concept of Value for the other stakeholder groups, and to implement appropriate Value Management processes.

References

ATSE (2016), A summary to the Report *Research Engagement for Australia. Measuring research engagement between universities and end users*. Available at https://www.atse.org.au/content/publications/reports/industry-innovation/research-engagement-for-australia.aspx.

Fisher, N I (2013), *Analytics for Leaders. A Performance Measurement System for Business Success*. Cambridge: Cambridge University Press.

Kordupleski, R. (2003), *Mastering Customer Value Management*. Pinnaflex Educational Resources, Inc.: Cincinnati, OH.

Tukey, John W. (1962), "The future of data analysis". *Annals of Mathematical Statistics* **33** (1), 1–67.