EVALUATION AND THE MEASUREMENT OF CULTURAL OUTCOMES

Te Kani Kingi

Te Pūmanawa Hauora/Te Mata o te Tau

Research School of Public Health

Massey University

WELLINGTON

Tauhara Centre, Taupo   **July 2005**

INTRODUCTION

Firstly, I would like to extend my thanks to the organising committee, and for the invitation to speak today. When initially approached to provide a paper at this conference it was with some enthusiasm that I accepted, especially given the overall theme of evaluation, and the focus on "leading by example" and "leading by association".

In considering what to present today, and how best to do this, I was faced with a familiar dilemma – and, with on the one hand constructing an entertaining and somewhat interesting presentation, while on the other offering something that was more informative and useful. Striking a balance here can be difficult in that we all appreciate the need to convey ideas in a vibrant and motivating ways, though are similarly aware that our subject matter or area of interest may not always permit this. Fortunately, the area of evaluation provides ample opportunity to explore a dynamic range of issues – to inform and enlighten, while at the same time engaging.

With these ideas in mind (and as a broad guide) I've decided to consider a number of quite different, though certainly related issues – historical, cultural, conceptual, and hopefully practical. And, while the overall intent is to offer a presentation that is both interesting and informative - as evaluators many of you will understand that the stated objectives and identified outcomes do not always match. In this regard, I'll let you be the judge.

A HISTORY OF MEASUREMENT

To begin with, I would like to acknowledge that there are of course a number of components to the process of evaluation, and like research, this process often depends on the nature and context of the evaluation, particular characteristics, profiles, and often personal dynamics. While evaluation frameworks provide useful tools, no two evaluations are ever completely the same, and it is therefore difficult to predict what problems might be encountered and what issues may arise.

Because of this, and while it is impossible to consider every aspect of evaluation, I wish to focus on the particular issue of measurement and the fundamental requirement measure or evaluate what is important, relevant, and significantly, that which is attributable to the service, provider, intervention, or activity. Indeed, this presentation should be qualified by the fact that it takes place within the overall context of service provision – and in particular health services as this is the area I am most familiar with.

As an introduction to this and keeping in mind the promise of an historical backdrop it is important to consider that such concerns over measurement, and in particular accuracy, are not new and that for many thousands of years humans have struggled with this issues. That is, how to measure what was observed and to do so in a consistent, reliable, and valid way. Space, distance, and time were often the focus of activity, and issues of this nature (like today) were typically assisted with the help of measurement tools or instruments.

As many of you would know *Archimedes* is often referred to as the father of modern mathematics and integral calculus. He was also a keen inventor and developed machines of war, planetariums, and quite possibly, the water screw. However, it was his interest in pycnometry – the measurement of the volume or density of an object – for which he is perhaps best known. In a quite famous, though apparently exaggerated story, it is said that when *Archimedes* got into his bath and saw it overflow, he then suddenly realised he could use water displacement to work out the volume and density of the king's crown. A task he was assigned when it appeared it had been constructed partly of silver, instead of gold. Archimedes not only shouted "Eureka" (I have found it) but he supposedly ran home naked in excitement.

While Archimedes interest in the measurement of area, density and volume revealed an advanced understanding of science and mathematics it also illustrated a more fundamental desire to apply knowledge in a practical and useful way – to measure something precisely, easily, and in a dependable way.

The measurement of time further reveals how we have struggled to quantify the passing days, seasons, and years and how open thinking, and technological advancements have allowed us to do so with greater precision. Celestial bodies, the sun, moon, and stars

were all used at some stage to consider the passing of time and also to navigate. We know that artifacts gathered from the prehistoric era show early attempts at time quantification. Bone and wood scratching from more than 20,000 years ago have been found, and thought to represent certain periods of time. Constructed some 4,000 years ago Stonehenge remains a modern mystery, however, the alignment of stones suggests its function in determining seasonal or celestial events, such as lunar eclipses, and solstices. Five thousand years ago, Sumerians in the Tigris-Euphrates valley (near present day Iraq) divided the year into thirty months and the day into 12 periods.[1]

The earliest Egyptian calendar was based on the moon's cycles, however, they realized that the "Dog Star" in Canis Major rose next to the sun every 365 days, about when the annual inundation of the Nile began. Based on this, they created a 365-day calendar in 4236 B.C - the earliest year ever recorded.

While it appears that at this point in time an annual base-line had been established, there was by no means a consistent or universal approach to the measurement of years – what number of months were appropriate, how many days should be included, and how each day could be partitioned. Days and nights were obvious measurement platforms as were the seasons – though, and depending on latitude, the length of both days and nights around the globe would vary, and in many countries seasonal variations in climate were only slight. In any event, and to a large extent, it didn't really matter anyway – as long as you knew when to plant, when to harvest, when to fish and hunt, there was no real need to more precisely quantify or measure time.

However, and about six thousand years ago a significant structural event took place (within society) and required more accurate means of daily timekeeping – that is bureaucracy. Sundials, obelisks and sun clocks are early example of how time (within each day) was accounted for. These were somewhat useful, though their effectiveness at night rather marginal. Mechanical water clocks were later developed, these offered even greater precision, worked just as well at night, though were cumbersome to construct and to use. Sometime later, spring and weight-powered clocks appeared as did pendulum clocks. With each advancement, clocks or time-peaces became more common place,

[1] http://inventors.about.com/library/weekly/aa070701a.htm

often more compact (though complex), but importantly more precise. The development of quartz based clocks in the 1930s and 40s offered a level of accuracy that had never before been possible – slipping a mere two seconds per month. Atomic clocks (first developed in 1948) slip just one second in a thousand years. In 1955, the first celestial beam clock was invented and is estimated to lose or gain just three seconds in a million years. The latest clocks are accurate to one second every 30,000,000 years.

While the examination of these issues and developments give some indication of both measurement and accuracy, they likewise raise four obvious questions. The first which springs to my mind is "why". That is, why do we need to measure things like time, space, or distance? Furthermore, why do we need to do so with greater precision? Certainly, these are of academic and scientific interest, but also have more pragmatic applications and often lead to advances elsewhere and in unexpected ways. Nevertheless, the "why" questions is all important as it provides purpose, direction, and overall focus. When attempting to measure volume and density, the why question for *Archimedes* was clear as was the purpose. As society has evolved so has the need to measure time in more exact ways and in order to inform various aspects of what we do.

This brings us to the next question, which is "how". Certainly, the examples given provide an understanding of this. However, of greater interest is the way in which the "how" question is ultimately answered, how it often spawns innovation and likewise reflects on issues of accuracy, reliability, validity, repeatability, and of course consistency.

The next question is "what" – that is, what are we attempting to measure. This is of course linked to the "why" question, though brings added focus to the fundamental purpose. For Archimedes the "what" question was density and volume, though conceivably could have been water displacement. Likewise the development of time pieces had as much to do with the passage of celestial bodies, movement, weight and the elasticity of metals, and more recently sub atomic oscillations. What is important therefore is that we have a clear understanding of what we are attempting to measure, and what we are not.

The final question is "when" – this is perhaps less obvious from the examples given, though considers the importance of both time and space. "When" reflects on the need to measure change, precisely and consistently, but importantly a point at which change is expected. In this regard there is little point in measuring anything unless some form of change – weight, density, or time is anticipated.

The examples given are used to illustrate a number of key points. The first being that issues of measurement are not new and is something that humans have been grappling with for thousands of years and within many different disciplines. We see also that with the passing of time advances and developments have occurred. Greater precision and accuracy, application of ideas and methods elsewhere, and an expanding knowledge base. However, it is the more fundamental questions that are perhaps of greatest interest to this conference and this paper, and which have also remained constant. That is, the fundamental requirement to consider the "why", the "how", the "what", and the "when".

THE WHY QUESTION

Insofar as evaluation is concerned the "why" question is by no means a simple one and often depends on your particular perspective. For evaluators this question may centre on academic, interest, or even financial considerations. For those being evaluated it may simply be a contractual requirement. For services users it may assist with decision making and in determining what meets their needs. For government it is likely to aid policy formulation, service description, funding allocation, scope or coverage.

Putting aside the very many perspectives of why evaluations should be conducted (at a service or provider level at least) my own views are somewhat more pragmatic. To this end the question can be seen as a means through which understanding is enhanced. Though simple enough this idea of "enhanced understanding" has a number of implications and obvious uses. An enhanced understanding can reveal whether or not something is working well or to capacity, what modifications are required, and what impediments exist. It should lead to improvements in design, regardless of how successful the initiative is, and identify potential pathways forward.

Policy makers often complain that more information is required and in order to effectively plan. Therefore, and if the purpose of evaluation is to enhance understanding, it should likewise guide planning at a higher level – so that the needs of people are met and in the most efficient and cost effective way. Needless to say the implications for service users are also apparent and consistent with their desire to make informed choices.

It is apparent, therefore, that a number of options are available when attempting to consider the "why" question. The notion of an enhanced understanding provides a platform for this – though regardless, the question of why an evaluation is being conducted is a fundamental one that must be asked before it takes place. Not only as an overall guide for the evaluation, but also to consider how the information will be used and to what end.

THE HOW QUESTION

While the "why" question can be considered at a philosophical level and peppered with more practical examples, the "how" question is less quantifiable for the simple reason that not all evaluations are the same and therefore require different information collection, interpretation, and application needs. I do not wish to delve too deeply into the "how" question, suffice to say that it very much depends on the nature of the evaluation. Both qualitative and quantitative measures/or indicators can be used and gathered from a range of different sources. These are often used to construct a picture and depending on the information obtained (or those that are used) can determine how this picture appears – how accurate or detailed it is.

Gathering the wrong information will mean that the picture described doesn't actually represent what it says it does. The wrong mix of information may mean that the picture is in black and white instead or colour. Incomplete gathering of information may lead to a picture that is out of focus or blurred, has parts missing, or is observed through the wrong lens. The biggest risk however is when a picture is interpreted to mean something it is not. The wrong information has been gathered, the picture is inaccurate, yet is accepted as true and precise. Good evaluators will of course attach caveats to any pictures they paint, identify what limitations exist and should also assist with the

interpretation of the picture – as with most pictures its meaning may not always be that obvious.

In any event, the how question is often concerned with the gathering of information, and, while it is impossible to describe what information is relevant in every particular situation, the range, accuracy, and interpretation of this is of seminal importance.

THE WHAT QUESTION

I have so far avoided discussion on issues of cultural interest and have instead focused on more generic concerns. This is deliberate in a sense and that I wanted to explore these in greater detail and within the "what" section of this presentation. It is when considering the issue of "what" to evaluate that I believe cultural factors are often missed. In the past cultural considerations within research and evaluation were more often than not connected to process or consultation issues. Who is conducting the evaluation? How are people involved? What mechanisms for feedback are possible? How are the results interpreted or described?

These are of course all important considerations and certainly assist with ensuring that the evaluation or research takes place within an appropriate context or framework – though does not necessarily guarantee accuracy or a positive outcome. However, the "what" issue is not always fully considered from a cultural perspective (at least in my experience) and therefore offers greater opportunity to consider it within this paper.

As with the issues previously described, concerns over what to consider are very much dependent on the nature and extent of the evaluation and there are of course numerous factors to reflect on. However, and very simply, what to measure/consider or evaluate is determined by the objectives of the activity and assessing how well this is done. As described previously various measures or indicators can be used to assist with this, though they must ultimately reflect what objectives are sought.

Measuring these objectives or outcomes has recently been the focus of research conducted at Massey University and in particular the identification of culturally specific outcomes in health. The basic premise of this work relates to the idea that while certain

service outcomes are generic to all providers (e.g. providing care that is safe, effective, cost efficient) there are perhaps outcomes that are more culturally aligned and therefore should feature when the outcomes of service effectiveness are evaluated.

The background to this work starts some 20 years ago and when the first Māori health services began to develop. Initially within mental health, these services were designed to provide care within a cultural context and underpinned by the notion that Māori would respond better to treatment if delivered in way that matched their cultural realities. For the most part, and initially at least, services included mechanisms through which cultural practices and protocol could be introduced within the treatment process. Though in time, more and more sophisticated models of service delivery developed.

The fundamental reason for including culture within a health service actually has little to do with culture itself, and, it is often mistakenly assumed that cultural programmes are aimed at cultural enhancement. This may in fact be an objective or rather process – however, the primary purpose is to improve health outcomes. In this regard culture has little place within a health service setting unless it improves health outcomes.

To a large degree cultural interventions are designed to do just that – and to enhance the expected health outcomes. A pōwhiri within a mental health service, for example, is essentially a process of encounter, a welcome, a greeting, and an exchange of ideas and minds. The process is obviously derived from Māori tradition or custom and has been used for many hundreds of years. However, the health objectives of this are sometimes lost in that a deeper analysis reveals that the whole process can also be quite settling, putting the consumer and their whānau at ease, providing reassurance, and creating an environment which supports recovery and rehabilitation. Consumers are often encouraged to play a formal role in this process, either as speakers or supporters. This is of course consistent with Māori custom, however it also recognises the range of skills that consumers possess and affirms the desire to ensure that they are not just idle participants within the process – but rather the focus.

Likewise, the introduction of Māori arts and crafts, and as part of therapy, has the effect of making the intervention more relevant and rewarding – encouraging ongoing participation. Te Reo Māori, Karakia, Kapa Haka, and Tikanga have also been used to

similar effect and with improved health outcomes the prime focus. In line with Mäori philosophies of health and well-being, Mäori models of health have also been used to guide service design, treatment interventions, and the outcomes that are sought. While it is often difficult to accurately describe the core components of a Mäori health service – certainly Mäori models of health (such as Te Whare Tapa Whā and Te Wheke) as well as this notion of holism is something that appears consistent.

These ideas in many ways shape the way in which Mäori services (and not just in health) are delivered. However, and more than this – they often influence what outcomes are identified and subsequently sought. Examples of the unique outcomes sought by Mäori services are not always obvious or documented, but rather assumed as part of a cultural approach to service delivery. Of significance to the field of evaluation is the idea that these outcomes contribute to the well-being of individuals and service effectiveness – though may not always be considered during the overall evaluation of service efficacy.

Sticking with mental health we also know that many of the existing measures of patient outcome, while useful, tend to focus on factors such as symptom ablation, compliance with medication, or level of psychosis. These are all important considerations and are measures that can be used to determine service efficacy and patient outcomes. However, these kind of measures may not match (entirely) the outcomes sought or outcomes produced by a Mäori mental health service – even though these other outcomes (that are not considered) are extremely vital to the health and well-being of Tangata Whaiora (Mäori mental health consumers).

Recent research reveals that the outcomes favored by both Mäori mental health services and consumers are much broader than what existing measurement tools are capable of considering. As noted, indicators like the absence or presence of symptoms are valuable measures of progress, however, for Mäori these need to be balanced against other, broader requirements – the capacity to communicate well with whänau, to participate within Mäori society, to be culturally and spiritually enriched, and to have physical health issues addressed.

Within other areas of health similar examples exist and which show how the purpose of an intervention can sometimes miss the point, as far as Māori are concerned, and when they fail to fully consider cultural perspectives or outcomes. I remember once asking a hypothetical series of questions to a kaumātua concerning the outcomes sought following a hip replacement. A standardized questionnaire was used and designed to be applied about six months following the operation and to determine whether or not mobility had improved or returned to an expected level. One of his responses to a certain question was particularly interesting. That is, when asked to comment on the question "as a result of the operation are you now more able to walk around the block" his answer was "I guess it makes sense – but, what do I want to walk around the block for, I've got my car for that"

While a somewhat simple retort, this comment reveals an important consideration, that is, the fundamental requirement to measure what is important to the consumer. In this regard a more relevant indicator of success would be to determine whether or not the surgery allowed the kaumātua to stand in one place for a given period of time. Although walking around the block may make little sense to him – the ability to stand in one place, on the marae, and during whaikorero is a much more relevant measure of progress and one that appreciates his cultural realities.

Therefore, and with regard to the "what" question, two particular issues are worth considering – at least from a Māori service perspective. The first is that generic indicators of efficacy should be considered as part of evaluation. These indicators can be identified through a review of relevant documentation, reporting systems, or contractual arrangements. Measures or indicators are likewise developed to reflect these. The point is that Māori providers are no less likely to value a perspective on whether or not key outputs or outcomes are being met and in identifying where enhancements can be made.

The second point however concerns the identification of outcomes or objectives and the fact that many services (particularly if they operate within a Māori paradigm) produce outcomes that are not always routinely measured or considered. As mentioned previously, the fact that these are not measured does not mean that they are any less important, or do not contribute to the fundamental objectives of the service. As a

consequence many services can rightly feel shortchanged in that the benefits of their activity are not always fully considered.

In my experience these additional or cultural outcomes are seldom referred to by a service and one suspects that they are almost taken for granted - in a sense. Nevertheless, they should at least be considered within the overall context of an evaluation. The obvious difficulty however is in the identification of these additional or culturally derived outcomes, being confident that they contribute to the service efficacy, as well as having some understanding of attribution and their relationship to service activity.

Providing guidance on what these additional or culturally aligned outcomes might be is difficult and obviously depends on the focus of the service, it's design, and what activities take place. Outcome questionnaires or tools can be used and to assist with this process – however, there are currently few tools designed to consider cultural outcomes. Again, the fundamental question concerns the identification of cultural outcomes – that positively contribute to the activities of the service, but which may not be considered as part of routine evaluation.

THE WHEN QUESTION

It is at this point I would like to introduce the "when" question. Unlike the previous "what" question, the issue of "when" is far easier to consider. And, quite simply it infers the need to measure or evaluate "when" an outcome from a service is expected. Of course, and depending on the nature of the evaluation, this may not always be possible – however, and if the purpose is to determine what outcomes from the service are produced and whether or not key objectives have been met – then timing is of critical importance.

The risks involved with getting the timing wrong can be considerable and are perhaps most usefully illustrated though examples from the health sector. In this regard it is unreasonable to expect immediate outcomes from care and often ongoing treatment will be required. A patient receiving counseling for an alcohol or drug problem is unlikely to show significant improvement after the first consultation and it is therefore unreasonable to evaluate the performance of the individual or program at a point when a positive

outcome is not yet expected. Likewise the success of a hip-replacement operation should not be undertaken as soon as the surgery is completed or even a few days later. A more reasonable approach would be to assess it after a full period of recovery. In a similar way health programmes – particularly public health initiatives, often take years before measurable outcomes can be expected. Encouraging individuals to take a more responsible approach to drinking, smoking, sex, or driving is something that is unlikely to generate immediate results. Yet, they often assessed in the short-term (by the media at least) and without being given the opportunity to run full-term.

Another related issue is the misconception with respect to what objectives or outcomes are possible. To this end, and only in rare cases, can we expect that the best possible outcome will be achieved. In fact negative outcomes are all too possible, and in some cases likely. Road safety campaigns are a good example, that if evaluated based on the numbers of deaths per year could suggest that the current strategies do not work – particularly if the numbers increase. However, this blunt measurement or indicator is perhaps too simplistic in that it can often take a generation for societal behaviors to change and even if they do – not everyone will. As well, an increase in the total number of deaths may in fact show that the programme is working – if the increase is lower than projected rates.

For all these reasons the "when" question is critically important and should be balanced against the need to set realistic outcomes – which in turn can be attributed to the particular activities of the service. The complexities which surround these types of questions and decision making processes are significant and it is not possible to offer a set of rigid or even generic guidelines. Suffice to say that it depends on context and situation – though must at least be considered.

PROXY OR OUTCOME

I have so far considered four key questions or issues which are derived somewhat from the historical examples given earlier on. In this regard early attempts to measure both space and time were governed by the need to also consider more pragmatic questions of "why" "what" "how" and "when". However, there is a fifth question or issue that is also worth considering in that in many cases "what" was being measured was not always that which was observed. For example; early attempts at measuring the passing of time such as hour glasses, water clocks, or sun dials were not actually measuring what they were designed to or purported to. More correctly they were instruments that considered the forces or gravity on water or sand as well as the relative position of the earth to the sun. In essence, they were proxy measures of time.

The use of proxy measures really depends on how accurate or representative they are. There is an obvious issue in that sometimes they don't actually measure what they are purported to, or used in ways they are not designed to or out of context. The real issue here is not whether they actually measure what they say they do (as in many cases this is not always possible) but how accurately they do this. In this regard it matters not that some of the most precise clocks in the world actually consider the relationships that exist between sub-atomic particles. Of more relevance is that fact that these reactions also/and by proxy, accurately consider the passing of time.

In terms of evaluation – proxy measures are often used to determine efficacy. While the fundamental purpose of health services are to improve the health of individuals or communities – often it is not possible to say with any certainty the extent to which this is happening or the overall role of the service. Often proxy measures are used in the absence of more accurate and reliable outcome measures. In this regard effectiveness is sometimes determined (for example) by the number of patients assessed, the number of vaccinations administered, the number of operations performed, or the number of times treatment was undertaken. Other proxy measures include the extent to which processes and procedures were followed, what accounting systems are in place, or how the service is managed. These are actually input, process, throughput, or output measures and don't actually consider whether or not the fundamental objectives of the service

(improvements in health) have been met. As a further illustration – there is frequently used medical aphorism which highlights this very point; and goes something like this;

The operation was a complete success.......unfortunately the patient died

While the example is fairly simple, it usefully illustrates the issue of using proxy measures. To this end, it shows that while the surgeon may have been well qualified, the equipment reliable and efficient, the support stuff available and competent, this does not necessarily guarantee that the outcome is positive (for the patient at least). Given these concerns over the use of proxy measures (such as inputs, outputs, or through-puts) you would probably guess that I am not a fan of these types of measures. However, this is not entirely true in that proxy measures can sometimes be extremely valuable – especially in the absence of more accurate information. Moreover, outcomes are unlikely unless certain inputs, outputs, and processes are also present.

The advantages are that proxy information is often more easily and routinely collected, can be more comparable, and not as difficult to identify or quantify. The more relevant issue is whether or not these proxies actually contribute to an understanding of effectiveness. If they do, then they should be considered when decisions on effectiveness are made – if they do not, then for obvious reasons they should be dispensed with. Again, the difficultly is when they are used to represent something they are not.

A good example of a situation where proxy measures should or can be used relates to vaccinations and the administration of these. As already noted, I described counting the number of people vaccinated or immunized as a proxy measure in that it doesn't actually consider whether or not and the extent to which people were protected. However, if the vaccine is proven and shown to work without major side-effects – then certainly one could have every confidence that a proxy measure (based on the number of vaccinations) is likely to reflect the actual outcome – improvements in health or protection from disease. In the same way, and in the absence of actual measures, proxies which consider programme management, activities, systems, or processes can likewise generate useful

information. Though again, should be used with some caution or an understanding of what caveat may be attached to them.

CONCLUSIONS

I have attempted within this presentation to raise some important considerations for the process of evaluation. This has been difficult in that it is impossible to draw absolute or definitive conclusions in that evaluation is often governed by context or situation, has a range of different meanings and does not always fit easily within a generic framework or box. In this regard, and depending on your experience or interest, parts of it are likely to be more useful than others.

The idea of taking what is useful and avoiding what is not is a skill that many evaluators have and in this regard are well adept at sifting through information, pulling out that which matters, and interpreting this in useful, accurate, and meaningful ways. The more experience you acquire, the easier this becomes, as key decisions are aided by past encounters, your own developing knowledge and skill base.

The issues described in this presentation should therefore be balanced against what you already know and likewise applied with discretion and depending on any number of contextual variables. However, and in recognition of this, it is important that some form of summary or concluding remarks are made and in order rationalize the interpretive process. Fortunately, the five major points made in this paper provide a reasonable framework for this.

The first point, therefore, is that broader philosophical or pragmatic questions will need to be asked and reflect on the actual purpose of what you are trying to achieve. This is of course a fairly obvious thing to do, though is not as ridiculous as it sounds. By asking this question before and during the process of evaluation certain parameters can be established and likewise an ongoing assessment is made as to whether or not key objectives are being met. As a young researcher I often fell into the trap of not having a clear understanding of "why" I was conducting a particular piece of work. Accordingly the fundamental purpose would often drift and result in the collection of information that, while somewhat interesting, had little to do with the actual investigation. Asking

the question as to "why" the work is being undertaken can also encourage discussion on how the information will be used, how it will contribute to the existing knowledge base and lead to positive developments.

The issue of "how" to measure was examined through a range of examples and qualified by the fact that this question really depends on the overall context and focus of the evaluation. However, it was stressed that whatever processes are adopted, the important thing was that these were accurate and robust, representative and legitimate. Likewise, that the information should be interpreted in a valid way.

The issue of "what" to measure provided the base through which cultural issues and considerations were explored. Two main points were made in this regard. The first was that Mäori services need to measured or evaluated against the more usual systemic and procedural indicators. In this regard they are no less likely to require an assessment of whether or not appropriate managerial or administrative systems are in place. Furthermore, the extent to which contractual arrangements have been met and the manner in which key objectives were completed. However, and given the way in which many Mäori services operate, it was likely that additional outcomes were produced, and, that these outcomes contributed to overall service effectiveness. To this end, it was suggested that a comprehensive evaluation of service efficacy must take into account these dual objectives. A failure to do so is unlikely to present a picture which is accurate or comprehensive.

The fourth point emphasised the need to get the timing right and that while activities and programmes are evaluated at different times and stages of development at the very least an assessment of time should take place. The real issue here is in the identification of appropriate outcomes, the fact that ideal outcomes are not always possible, as well as the idea that a developmental period is required. The best possible outcomes may take years to manifest, so in the meantime more realistic objectives are required.

The final point concerns the use of proxy measures. For good reason, many avoid the use of proxies in that, and by their very nature, they do not actually measure what they are used for. However, and from my point of view, their application has little to do with whether or not they measure this or that – but how accurate or close to a proxy they are.

In this regard, the gathering, application, and interpretation of a range of proxy measures can prove to be very useful in the absence of more focused information. The key therefore, is to collect the right kind and range of proxy information and being confident that this accurately represents what you are really wanting to investigate.

The framework below is a basic summary of the key messages from this presentation. While I don't expect that it will cause you to suddenly run from this room and shout "eureka" - I do at least hope that you have found it useful and interesting, but at the very least a platform for future discussion. Kia ora.

Why ?	Offers purpose and direction, parameters and focus
How ?	Depends on context and situation, but must be as robust and accurate possible
What ?	Generic vs Cultural indicators. Measuring all of what is potentia produced
When ?	Consideration of when outcomes are expected and what they might be
Proxy ?	Being sure of what you collect reflects what you are measuring