Can Evaluation Be Too Prescriptive?
Appraisal in the Age of the Triple Bottom Line

Ian Ker

Ian Ker
iank@arrb.com.au
ARRB Transport Research

Paper presented at the 2002 Australasian Evaluation Society International Conference, October/November 2002 - Wollongong, Australia. http://www.aes.asn.au

Abstract

Simple evaluation frameworks lend themselves to clear specification. As governments and others increasingly recognise the interconnectedness of initiatives in achieving a wide range of outcomes (the so-called 'triple bottom line') and build these broad outcomes into the measurement of performance, the task of the evaluator becomes ever more complex. Complex evaluation requires more flexible frameworks, analysis and presentation so that information is exposed rather than concelaed and that judgements are made transparently and by those who have been entrusted with the responsibility of doing so. Indeed, we can no longer consider the process and content of evaluation in isolation from its presentation.

Conventional socio-economic and multi-criteria frameworks, for so long the staple of public sector evaluation in areas such as infrastructure investment and market-based service provision (such as public transport) have consistently had difficulty in addressing the triple bottom line, usually relegating social and environmental impacts to the twilight zone of 'externalities' and resorting to sensitivity analysis where judgements are made.

This paper discusses issues raised by the triple bottom line requirements, how they are being addressed in evaluation practice, including some specific examples, and what further developments are on the horizon.

Key Words

evaluation, appraisal, benefit-cost, performance measurement, transport, triple bottom line, goals achievement

Introduction

It is a common-place to say that assessment and related ex-post performance measurement is difficult, even impossible, for policy areas of the public sector. After all, there are so many factors that can affect the implementation and effectiveness of policy – and by implication, these are always adverse impacts, so the outcomes will be less favourable than we had stated, and 'political', so beyond the 'control' of the public servant.

Less commonly do we see reference to any of the reasons closer to home why such ex-ante assessment or ex-post measurement rarely lives up to the rhetoric. These might include:

- being unclear about the outcomes we are trying to achieve;
- inadequate understanding of the complex systems within which policy must operate;
- inadequate recognition of the potential for unintended consequences;
- over-analysis, or its companion, over-reductionism – studying the parts in ever-greater detail rather than understanding the whole;
- attempting to optimise for specific future(s) rather than seeking robustness;
- over-stating potential impacts in an attempt to maximise the likelihood of policy being adopted.

A better understanding, or at least acknowledgment, of these might lead us to a more realistic view of assessment and performance measurement in the world of policy.

The World of Transport Policy

Transport strategies have changed direction very substantially since Goodwin (1990) coined the phrase *The New Realism*, but program delivery and the methodology of evaluation have not kept up, often because the linkages between new initiatives and outcomes are not clearly-enough defined or well-enough quantified. In addition, evaluation methodologies often assume that 'more is better' and have difficulty coping with change that includes changes in what we do (activity patterns) as well as how we get there (travel). Consequently, new initiatives often have great difficulty getting funding.

Put another way, we commonly set targets for a reduction in the level of reliance on private car travel in cities, but do little or nothing to increase the likelihood of those outcomes being achieved.

In addition to State Transport Policies and Metropolitan Transport Strategies embodying both the direction and the magnitudes of these targets, we have implementation strategies to tell us at an almost operational level what needs to be done.

Since many of these initiatives are new, we do not know, a-priori, how effective they will be. This poses problems for both the initial evaluation and for post-implementation monitoring. Ironically, lack of experience from which to judge likely effectiveness makes it all the more important to be able to measure actual performance.

It is increasingly recognised that transport systems and cities are complex adaptive systems, not mechanisms with uniquely-definable relationships between actions and outcomes. Complex systems have two key characteristics that make performance measurement difficult:

- they cannot be controlled only influenced or disturbed – the precise outcome cannot be identified a-priori; and
- 'emergence' of new characteristics, defined as outcomes that cannot be identified, let alone quantified, from our understanding of the system (Chambers and Ker, 1997).

The World of Transport

It is a fundamental of transport economics that the demand for transport is a derived demand. In other words, transport and travel are not valued for their own sakes but for what they enable us to achieve in other respects – economic, social and environmental. That being so, there are circumstances in which we seek to minimise the amount of transport (to achieve given outcomes) as well as occasions when 'more is better' (if the range of outcomes is increased). The common practice of evaluating transport initiatives in purely transport terms (such as travel time or vehicle operating costs) does not provide the basis for distinguishing between these circumstances.

The economics of the two situations are very different. In the former case (Figure 1), the change in the amount of resources required for transport is, in most cases, an adequate measure of the benefit from the initiative (but see SACTRA, 1999, for circumstances in which this might not be the case).

In the latter case (Figure 2), the benefit is reduced by the consequences of the increase in the level of demand – especially where there is congestion in the system (ie the supply curve slopes steeply upwards), in which case the benefit can disappear (Ker, 1989). Since it is precisely where there is congestion that the phenomenon of induced demand is most likely to be found and where the apparent advantages of adding capacity are greatest, it is not surprising that there is demonstrated evidence of significant induced travel (SACTRA, 1994) from major road projects.

It is perhaps a little less expected (but nonetheless logical) that the converse also applies – reductions in road capacity can actually cause some traffic to disappear from the system (Cairns, et al, 1998).

Dislocated Process

It is difficult for transport professionals (and professionals in other disciplines) to acknowledge that unforeseeable impacts might be significant. It is even more difficult to acknowledge that they might be

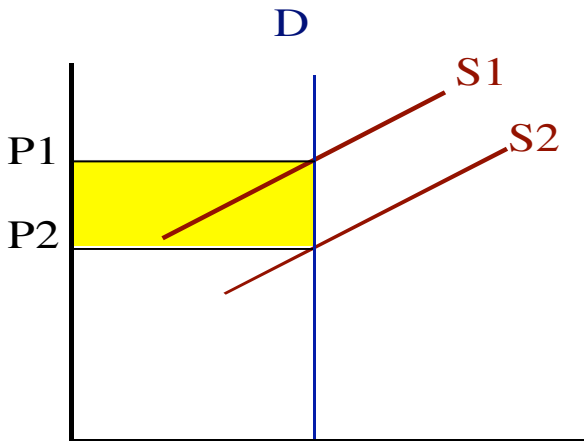

Figure 1. Conventional Evaluation

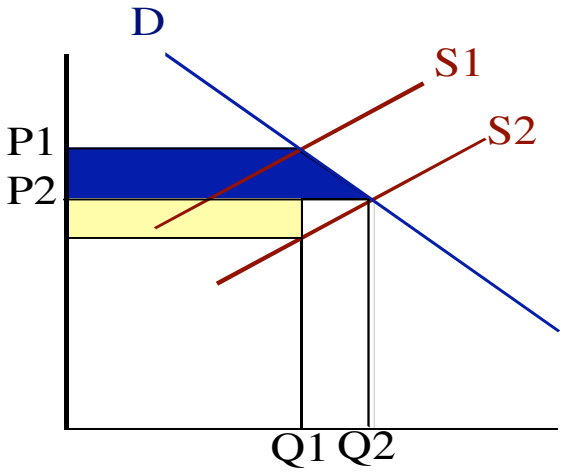

Figure 2. Evaluation with Induced Demand

intrinsically unmeasurable – so we measure what can be easily measured (activity) rather than what is important (outcomes).

This is usually reflected in a dislocation in process (Figure 3), characterised by:

- development of policy and strategy that has regard to the purposes the community wishes to have served by transport – of ten expressed in terms of 'access' or 'accessibility', which *is the ease of obtaining goods or the benefits of an activity (work, recreation, education, shopping, medical services, etc). Accessibility can be achieved through mobility, proximity or the use of electronic communications.* (Transport, 1995)
- network and project planning and implementation that has regard to the efficiency of the system – usually expressed, in transport, in terms of 'mobility'. This has the added complication that the models used for planning start from the basis of 'trips', which have, at least in principle, a direct link to activity, but the only output that is measured is generally movement (vehicle traffic).

When we try to measure performance in transport there is rarely agreement on <u>what</u> should be measured, never mind <u>how</u> it should measured.

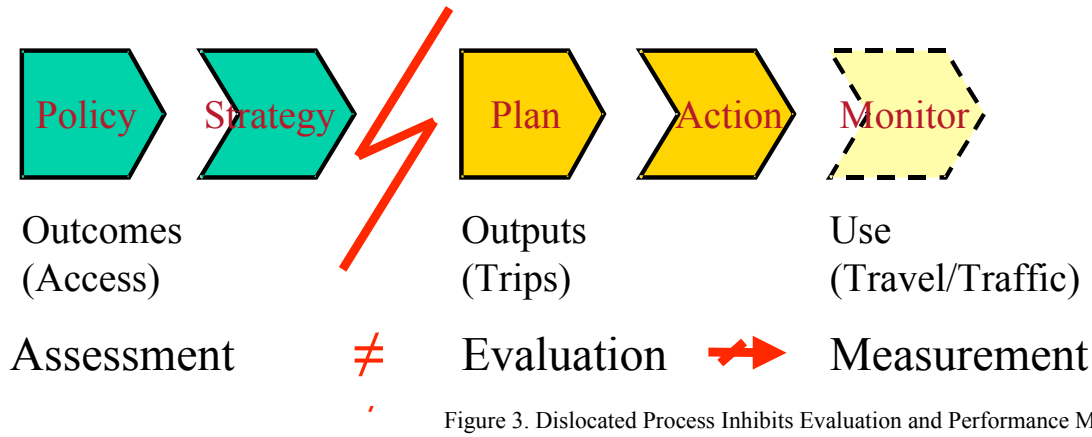

Figure 3. Dislocated Process Inhibits Evaluation and Performance Measurement

Unintended Consequences

The conundrum of better transport leading to worse services is widely recognised, especially in rural Australia (Figure 4), where the loss of services and population can be directly traced to improvement in transport systems, compounded by a narrow focus for both public and private sector decision-making in such areas as education, health services and banking. Decisions in these (and other sectors) have been based on 'privatising' the benefits of improved transport and transferring transport costs to the consumer – who has to travel further to access schools, hospitals and financial services.

This is not to say that there have not been benefits of concentration in these activities, but the continued complaints about lack (and loss) of service in rural Australia strongly suggest that the benefits do not come anywhere near compensating for the costs (including inconvenience) to the consumer.

Some years ago, *The Bulletin* ran a major feature on the depopulation of the bush that made the same points. The ultimate expression is the 'fly-in/fly-out' workforce in parts of the mining industry, which ensures that most of the income generated through the mining is spent out of the area and contributes little or nothing to the development of communities.

Increasingly, the same cry is being heard in the cities as well, as commercial rationalisation decisions in these same sectors lead not only to longer journeys but also to systematic discrimination against those who do not fit the planning paradigm of universal car access – including young people, the aged and people with disabilities. Interestingly, the planning paradigm of universal car access does not apply to more than a bare majority of the population, as up to 45% of people either do not have a driver's licence or do not have access to a car at any particular time. Add to that the fact that those with access to a car will need to act as chauffeur for the 45% with increasing frequency and the inconvenience that results form the dominant planning paradigm starts to affect the majority.

Ultimately, we have to admit that decision-making based on conventional evaluation has not always delivered unambiguous benefits in terms of what the community really values (Eckersley, 1997).

Making the Links

In principle, the solution is simple – establish a common metric and analysis framework for all three stages.

In practice, because planning is seen to require more detailed estimates of demand and usage, planning models do not provide information on a comparable basis to strategic assessments; few attempts have been made to relate the results of traffic models to measures of access. Although there are a few accessibility models, they are not part of the mainstream approach and are rarely used in practice.

The key questions, however, should be ones of robustness, ex-ante, and relevance, ex-post, rather than one of often-spurious precision.

Triple Bottom Lines in Evaluation

There is nothing new in multiple economic, social and environmental objectives for transport projects. For example, major road projects will often state objectives in respect of:

- transport - such as reducing traffic in congested areas, improving access by other means, reducing traffic on local streets and reducing road trauma;
- land use planning - such as minimising planning blight, providing a catalyst for redevelopment and maintaining economic activity; and
- social, environmental and economic - such as encouraging inner city living, heritage retention, enhancing amenity, improving equity of access and minimising air and noise pollution (Burswood, 1993).

The evaluation response, however, has usually been poor in terms of measuring those impacts that are not direct transport ones such as travel time and travel costs. Moreover, even where broader impacts are quantified, the evaluator usually attempts to reduce these to a single cardinal or ordinal measure, either in:

- benefit-cost analysis, in which all things are reduced to a monetary equivalent; or
- multi-criteria analysis, in which a range of criteria are weighted to produce a ranking of projects for decision-making purposes.

What is rarely acknowledged is that these two methods are mathematically identical, since both require a common unit of measure to compare unlike impacts. Whilst there has often been criticism of benefit-cost analysis on the basis of its reducing everything to money, the most commonly-used alternative is in essence no different.

The real problem with both approaches, in the age of the triple bottom line and whole-of-government performance measurement is that a very large amount of information is lost in the process of measurement and aggregation.

The alterative of some form of Goals Achievement Matrix is rarely considered by technocrats on the grounds that it is too difficult to interpret and does not give unique and clear results. In the United Kingdom, the

development of a 'New Approach to Appraisal' (DETR, 1998) has included the development of an Assessment Summary Table (AST) as the core of the integrated assessment process - this is effectively a goals achievement matrix. What is more, this form of disaggregated presentation of impacts has been demonstrated to have a systematic and transparent influence on decision-making (Nellthorp and Mackie, 2000).

A modified AST is under development in Western Australia (Figure 5).

PROJECT TITLE:	Enter Title Here	**PROJECT NO:**	02/xxx
PROBLEM/OBJECTIVE	What is the project/program intended to achieve?		
KEY DRIVERS			
Other options considered	List other options that have been or could be tested (independently or in conjunction with this one) and why they are regarded as less satisfactory than this one		
Status/Outstanding Issues	How ready for implementation? What issues still to be addressed?		
Policies/Strategies supported: Transport	Title(s) and specific reference(s) (eg Metropolitan Transport Strategy – Walking; Perth walking – Promote safe and secure walking environments)		
Policies/Strategies supported: Other	Title(s) and specific reference(s) (eg Sustainability Strategy – reducing car dependence)		

Part 1: Strategic Assessment

STRATEGIC DIRECTION	OBJECTIVE/OUTCOME	QUALITATIVE IMPACTS	QUANTITATIVE IMPACTS	ASSESSMENT
Integration	Impacts on whole-of-government performance measures			
	Intermodalism			
Safety	Fatalities			
	Injuries			
	Personal security			
Accessibility	Pedestrians and cyclists			
	People with disabilities			
	Access to public transport			
	Affordability			
Sustainability	Greenhouse			
	Air Quality			
	Energy			
	Congestion			
	Regional Development			

Figure 5 Appraisal Summary Table (Criteria and content illustrative only)

An Urban Transport Investment Assessment Framework

In 1998, the national taskforce on Measure 5.3 of the National Greenhouse Strategy, 'Promoting Best Practice in Transport and Land Use Planning', posed the question of what an integrated assessment framework (Figure 6) for urban transport investment, taking into account greenhouse issues, would look like and how it might be implemented. The outcome challenged some of the key assumptions of traditional transport evaluations and highlighted some key areas of improvement necessary if evaluation were to play an effective role in transport decision-making. The most significant of these related to:

- The identification of relevant options to evaluate;
- Understanding and estimating the cause-effect relationship of interventions;
- Allowing for feedback;
- Including all significant outcomes in the evaluation; and
- exposing rather than hiding component information on costs and benefits (Allen Consulting, 2000).

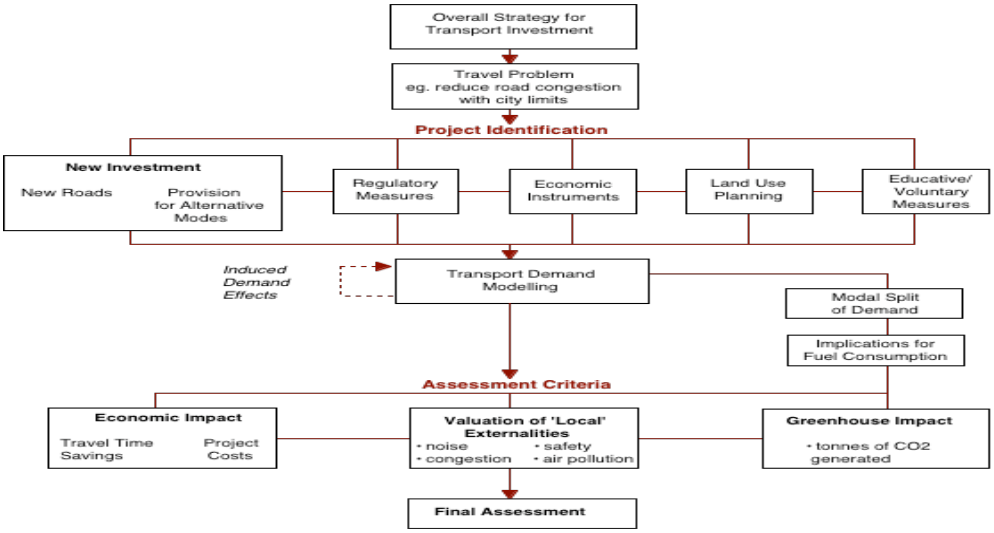

Figure 6 Integrated Assessment Framework for Urban Transport

Double-Counting - Does It Matter?

Traditional socio-economic evaluation methods have consistently resisted the tendency to double-count benefits. For example, it is well-established that travel time savings from urban transport improvements are to a large extent capitalised into property values, so we count only the travel time savings and regard the property value impacts as a distributional issue.

This is entirely correct when we are attempting to encapsulate evaluation in a single value. But what if we are specifically interested in the distribution of benefits, as we are when addressing the triple bottom line? In this case, we must clearly present all impacts in a disaggregate form, provided we make it clear where the benefits end up. Thus, it is important we demonstrate where travel time savings end up (for example, they may be taken out in travelling further for the same activities rather than in travelling the same amount in less time and using the time for something other than travel[1]).

Evaluating New Directions

Transport strategies have changed direction very substantially in the past decade or so, but the methodology of evaluation has not kept up, often because the linkages between new initiatives and outcomes are not clearly-enough defined. In addition, evaluation often implicitly assumes that more is better and have difficulty coping with changes in what we do (activity patterns) as well as how we get there (travel). Consequently, new initiatives in transport often have great difficulty in getting funding (Ker, 2001).

Demonstrating and evaluating the effectiveness of transport programs depends critically upon being able to determine whether the outcome changed significantly and whether the program or something else caused the change (Higgins and Johnson, 1999). Both of these issues are especially problematic in the context of travel demand management and pedestrian strategies.

Travel Demand Management

Demand management is a relatively new concept in transport and until recently was seen almost entirely in supply-side terms. However, initiatives in Perth and Adelaide have successfully changed people's travel behaviour through 'soft' measures using information, opportunity and incentive aimed at individuals and households - TravelSmart Individualised Marketing (TSIM - Perth) and Travel Blending (Adelaide).

Both TSIM and Travel Blending proceeded through pilot projects which were then evaluated using conventional benefit-cost frameworks. The South Perth pilot of TSIM was constructed according to a rigorous experimental design, to ensure that the true effectiveness of the intervention were being measured, uninfluenced by extraneous factors - including publicity. Whilst conventional benefit-cost measures such as benefit-cost ratio and net present value were calculated, the evaluation included clear statements of the component impacts (Ker and James, 2000).

More recently, the evaluation has been enhanced to include specific estimation of a wider range of impacts (Ker, 2002), including:

- health and fitness; and
- public sector, cross-sectoral financial impacts (mainly in the health sector).

Walking as Transport

Hillman (2001, p42) draws attention to the need to 'broaden the aspects [taken] into account in [the] appraisal process for determining the most cost-effective strategies [for] transport investment … [incorporating] the [ublic benefits of promoting health through encouraging walking and improving the environment at the local and global levels by reducing air pollution and greenhouse gas emissions'.

The key problem with applying this to walking strategies, however, is the lack of knowledge about how effective any specific strategy is likely to be.

Using the evaluation framework developed for TSIM, Ker (2001) developed robust estimates of the benefits that would accrue from the achievement of the objectives of the Perth Walking strategy (Transport 2000). This provides the information necessary for decision-makers to assess how much achieving the objectives of the strategy is 'worth'. The focus is then clearly on the project proponents to demonstrate that the objectives can be achieved!

---

[1] The average travel time per person per day is consistently 55-65 minutes, irrespective of the type of city, predominant mode of transport or average speed of travel, for the same number of activities.

Conclusion

There is more than one way to skin a cat - but we have to know that it is a cat, why we want to skin it and what we want to do with the skin. Clear statements of objectives, models that are designed to measure against them and straightforward disaggregated presentation of evaluation measures are essential.

References

Allen Consulting (2000). *Urban Transport and Climate Change: Developing and Integrated Investment Framework*. The Allen Consulting Group FOR Transport WA, Australian Greenhouse Office, Transport Queensland and Victorian Department of Infrastructure: Perth, WA.

Burswood (1993). *Burswood Bridge and Road: City Bypass and Access Study*. Burswod Bridge and Road Committee FOR Department of Planning and Urban Development: Perth, WA.

Cairns, Sally, Hass-Klau, Carmen & Goodwin, Phil (1998). *Traffic Impact of Highway Capacity Reductions: Assessment of the Evidence*. Landor Publishing: London, UK.

Chambers, L C and Ker, I R (1997). Paradigm Shifts in Transport: Today's Heresy/Tomorrow's Wisdom. *Proceedings 21st Australasian Transport Research Forum*, Adelaide, South Australia.

Eckersley, Richard (1997). *Perspectives on Progress: is life getting better?* Working Paper 97/27, Resource Futures Program, CSIRO Wildlife And Ecology.

Higgins, T J and Johnson, W L (1999). Evaluating Transport Programs: neglected principles. Transportation 26, 326-336.

Hillman, M (2001). Prioritising Policy and Practice to Favour Walking. *World Transport Policy and Practice*, 7 (4), 39-43.

Ker, I R (1989). Putting the Economics back into Evaluation: Insights into the Evaluation of Urban Road Projects. *Proc. 14th Australasian Transport Research Forum*, Perth, WA

Ker, I R (2001). Deconstructing the Future: Assessing new initiatives in transport, including demand management and walking. *World Transport Policy and Practice*, 7 (4), 55-60.

Ker, I R (2002). *Preliminary Evaluation of the Financial Impacts and Outcomes of the TravelSmart Individualised Marketing Program*. ARRB Transport Research FOR WA Planning and Infrastructure: Perth, WA.

Ker, I R and James, B (2000). *The Benefits of Driving Less: Evalaution of the TravelSmart pilot program in South Perth*. Report 396, Transport WA: Perth, WA.

Nellthorp, J and Mackie, P J (2000). The UK Roads Review - a hedonic model of decision-making. *Transport Policy*, 7(1), pp 127-138.

SACTRA (1999). *Transport and the Economy*. Standing Advisory Committee on Trunk Road Assessment. HMSO: London, England. http://www.roads.dft.gov.uk/roadnetwork/sactra/report99/index.htm

Transport (1995). *Metropolitan Transport Strategy*. Transport WA: Perth.

Transport (2000). *Perth Walking: The Metropolitan Region Pedestrian Strategy*. Transport WA: Perth, WA.