BENCHMARKING PUBLIC TRANSPORT TO IMPROVE PERFORMANCE

D. P. Ashmore, F. A. Beltrandi and A Mellor[1]
Business Strategy Group
Steer Davies Gleave
United Kingdom

Abstract

The term "benchmarking" is often glibly used instead of "comparing". This fails to do the technique justice - a benchmarking exercise is far richer and more practical than a simple comparison. In the field of public transport (and other industries for that matter) it represents an essential tool to ascertain best practice in service delivery, and transfer these methods to other systems to improve their performance. There are several stages to a benchmarking exercise - the development of key performance indicators, the collection and validation of data to formulate these indicators, case studies to ascertain best practice, and the transfer of that best practice to appropriate environments. Comparison of data is merely a subset within the process.

In an era where public transportation systems are coming under pressure to attract people away from cars to ease congestion and pollution, and yet to do this with lower levels of public sector support, the need to learn from other systems is crucial. In addition in the United Kingdom, the drive for best value in Local Government will require insights into how others elsewhere do things, to what effect, and if similar methods could be employed in the UK. The authors illustrate the classical steps in a public transport benchmarking exercise using real life case studies developed in their professional activities around the globe - Argentina, Slovakia and several European urban operators. They then demonstrate how this methodology could be applied to facilitate overall public transport improvements. The paper therefore is designed not to be purely theoretical, but to illustrate a practical tool by which those in the transport arena can apply to meet their public and business objectives.

1 INTRODUCTION

This paper presents the fundamental principles of a management technique known as benchmarking, a technique that has been used with great success in several industries including transportation. In truth the technique is applicable anywhere as it is underwritten by very simple and portable principles.

The paper is structured in the following way. Section 2 offers an introduction to the technique and Section 3 describes the steps inherent in a benchmarking exercise, and demonstrates how it can be applied in the transport sector. Section 4-6 illustrates the technique using case studies from Eastern and Central European

[1] The Authors are Principal Consultant and Director of Steer Davies Gleave's Business Strategy Group in London and can be contacted on +44 207 919 8500 or at d.ashmore@sdgworld.net and f.beltrandi@sdgworld.net .

© Association for European Transport 2001

Railways, UK airport surface access studies and finally urban metro systems. Section 7 concludes the paper.

2 BENCHMARKING – THE BASICS

"If you know the enemy and know yourself, you need not fear the result of a hundred battles" - Sun Tzu 500 BC.

Most management techniques have their day in the sun, and benchmarking is no exception. It is fair to say that the technique is very much in vogue and has become one of the business buzzwords of the current era (although as the above quote from Imperial China demonstrates it has been in use for considerably longer). As such it has been hijacked by many so called "experts" who present it as some form of esoteric and complicated process. Nothing in fact could be further from the truth. Benchmarking is an extremely simple technique and one that the vast majority of us use in everyday life. Quite simply:

Benchmarking involves comparing your own performance to those of others undertaking the same, or a broadly comparable activity, to see why they do things better or worse than you. This will then allow you to do things better yourself or ensure you continue to do better by avoiding their mistakes, if your environment allows you to do so.

It really is that simple. If ever the reader becomes confused once the terms "Key Performance Indicators" (KPIs), "Case Studies" and "Best Practice" are used, he or she should refer back to this definition for comfort.

Before Transportation is even mentioned a simple example will illustrate the point. Suppose a person is concerned that he or she can never afford a holiday, whereas peers on similar salaries have several a year. An examination of the outgoings of his/her colleagues, however, demonstrates that they lead significant less lavish lifestyles with lower monthly outgoings. The person therefore reigns in his or her spending by whatever means necessary, putting the surplus into a separate savings account, and once the fruits of this become apparent, takes a well-earned holiday.

The person in question, although they probably don't know it, has carried out a benchmarking exercise. Where businesses are concerned, however, the number of variables and relationships can be much more analytically complicated and entwined, to make the task possible without some form of analysis that uses the formal steps of a classical benchmarking exercise (and employed in the above simple case study. These classic steps will be described in the next section.

3 STEPS IN A BENCHMARKING EXERCISE

The classical benchmarking exercise uses the following steps:

a. Formulation of the group of businesses/ services to be benchmarked
b. Analysis of the operating environment
c. Formulation of key performance indicators
d. Development of case studies to ascertain best practice for a given environment
e. Implementation of best practice to improve performance
f. Begin the process again – benchmarking is a continuous process

Steps a – e will be described in turn.

© Association for European Transport 2001

3.1 Formulation of the group of businesses/ services to be benchmarked

Businesses or services don't simply opt into a benchmarking consortium with no clear need. Often a business is struggling against its competitors and wants to benchmark against them; obviously the competitors will not be so keen to help them out. In this instance then the company will have plenty of data on itself (or maybe not and that would be indicative of a problem) but it would need to use published data for its competitors. Another example of where published data would need to be used would be when a consultant or analyst is undertaking a benchmarking analysis on his or her initiative without the knowledge of the parties being benchmarked.

The only instance when benchmarking clubs can be formed is when the services or businesses are not competing. Even in a global business context this would be unusual. Public services in different countries are therefore most likely to form a benchmarking club.

Once the club has been formed and the analytical team appointed the next stage will be an analysis of the operating environment.

3.2 Analysis of the operating environment

It is pointless comparing the performance of businesses or services unless you have at least a very approximate idea of the environment in which the business/ service operates, and the business/ service itself. For this reason the first stage of any benchmarking exercise should be an on-the-ground review, appraising things such as:

- An overview of the demand served
- The supply bases – assets, age and functionality.
- Labour base
- Political environment
- Profitability/ funding streams/ subsidies

An understanding of these issues will help to provide an initial understanding of why performance may differ. Once this review has been completed and documented the next phase of the work is to develop the key performance indicator framework.

3.3 Formulation of key performance indicators

You then need to compare the performance of each member of the group against the others by formulating measurements of comparison known as key performance indicators, henceforth known as KPIs. Within this activity there are several steps:

- Decide which areas you are seeking to compare
- Develop indicators for comparing these areas
- Develop very rigid definitions
- Compile a datasheet
- Send out the datasheet to the group
- Once sent back, validate the data, and resend to each member with queries
- Once data is validated, compile the key performance indicator graphs

Each of these will be discussed below:

- Decide which areas you are seeking to compare and what indicators you would

© Association for European Transport 2001

use to compare them. Suggestions could include:

- *Financial effectiveness - cost coverage* (operating/ investment) - how well do we invest in our system, control costs and maximise revenues? A good case study here would be investment in new technology or assets to improve reliability – do people do it and does it have an effect? To compare financial elements in the public sector a clear understanding of funding streams is needed, i.e. does the money come from the Government or can the business borrow on the open market. Purchasing power parity is an obvious concern – the value of a currency is important relative to what you can buy with a set amount.
- *Efficiency* - what do we get out for what we put in? In public transport you typically get out things such as vehicle kms or passengers journeys having put money, materials/systems and people in. For this reason this measure is usually used to show labor efficiency.
- *Reliability/service quality* - How much can the service be relied upon? How "good" is the service compared to others?
- *Asset utilisation?* How much do we "sweat" our assets in terms of infrastructure loading and passengers?
- *Safety.* How many people are killed or injured as a result of using our service?

- Develop indicators for comparing the areas of performance decided upon. Having decided upon what you want to look at, the next key question is how can it be measured and compared? This is where benchmarking gets quantitative. In fact rather than picking and choosing our measures it is often a case of choosing what we can measure with data in the possession of all the parties unless significant time is to be invested in new data collection.

  For each category it is often better to have a nest of indicators. For example some networks are in very dense urban areas with little competition so ridership will be higher there than for others. Others may have older infrastructure so that maintenance costs will inevitably be higher. Systems operating lengthy services far out into the suburbs may feel aggrieved at their output measure being normalised by route kilometers, because they have little leverage over the improvement of one element of the indicator. For this reason it makes sense to normalise by a variety of indicators - passenger journeys, vehicle kms, passenger kilometres, capacity kilometres.

  For illustrative purposes *suggested* indicators for the areas identified above could include:

  - *Financial effectiveness* - operating costs covered by fare revenue. This brings into play issues relating to fare level, ridership, labour costs, fuel, maintenance cost, administration , subsidy and federal grants.
  - *Reliability* - vehicle kilometres or passenger journeys per incident, or adherence to timetable or percentage of headways missed.
  - *Safety* - deaths or serious injuries per vehicle km.
  - *Asset utilisation* - capacity kilometres per passenger kilometre or percentage of vehicles spaces occupied in both the peak and the off peak. The former one shows the utilisation over the whole of the journey.
  - *Efficiency* - staff hours or investment capital per measure of output - vehicle kms or passenger journeys.

There will normally be problems with data availability, definitions and different

© Association for European Transport 2001

collection methods. Unfortunately as with many things the richness of the comparison graphs produced are a function of the weakest data in the group because all of the others often have to be downgraded to enable comparison. One useful by-product of this, however, is that the business/service with the worst performance (invariably those with the worst data), get to see how the more successful businesses measure their performance. They can thus modify their data collection systems as necessary.

- Develop realistic rigid definitions

To prevent the comparison of completely different items of data there must be a definition for each item. Otherwise you will get back what is known in the trade as "apples and oranges". These data definitions should be as sophisticated as is practical – the principle is that rudimentary data is better than none. If the definition is in-line with the most vigorous member of the group then often the others will not be able to deliver. This is especially true for reliability in the provision of service, where some may only measure cancelled services, whereas others may go into detail as to the fluctuation from the timetable and the evenness of the headways.

- Compile a datasheet

This should be as succinct and as user - friendly as possible with definitions next to the item. It should be done in a spreadsheet to enable direct and usable data entry.

- Send out the datasheet to the group

One of the most important attributes of success for a benchmarking exercise is that the exercise is supported through the entire organisation, and the results implemented at the highest level. It is unreasonable though to expect the managing director to go around an organisation collecting data. This task is nearly always delegated to someone junior - very often the quality of this person is crucial to the success of the exercise. To begin with they must be empowered from the highest level, otherwise it is not unusual for them to be treated in a dismissive manner by other members of the organization and given whatever data is available - often not compatible with provided definitions in the datasheet. It is also preferable for the benchmarking analyst to liaise with the data collectors, preferably on the ground so that they can be helped to understand exactly what is required of them.

A minor but useful point is that it is preferable to send out and receive the datasheet via email, to not only save time on data entry, but also to enable faster validation.

- Once sent back, validate the data, and resend to each member with queries

Here the analyst needs to undertake some sense checks on the data to make sure that there are no obvious gaping errors – for example if we are looking at a rail system, if the car kms divided by the car hours do not yield something sensible for average speed for the entire journey then one of the data items is wrong. In a large benchmarking exercise some datasheets will end up with people who simply put whatever is to hand into the datasheet, often from different sources, and send it back to the analyst. Data will often be unsourced and will not adhere to the definitions. There thus often begins a frustrating and time consuming set of iterations to ensure that the data is fit to go into the KPI formulation and comparison stage.

© Association for European Transport 2001

- Once data validated, compile the key performance indicator graphs

Having collated the best possible dataset for parties within the timeframe, the analyst then constructs the key performance indicator graphs. These can be simple bar charts, but to make them less dull for the reader, rudimentary hypotheses can be investigated by means of regressions and secondary axes. The graphs must be easy to understand, with key points annotated, although the temptation to draw conclusions at this stage should be resisted. Having shown how performance differs the next stage (the case study) ask the question why and whether or not methods are transferable.

To conclude this section it should be fairly obvious that the compilation of key performance indicators is labor intensive and time consuming. Despite it not being the most exciting part of benchmarking studies for most people, it should not be rushed – all too often a week is set aside for this phase – this is not realistic and the results will more than likely be flawed.

3.4 Development of case studies to ascertain best practice for a given environment

Hopefully the previous section will not have dissuaded anyone, that benchmarking is a simple process. All that it purports to do is to show that in order to compare business and services, you need to decide upon what you want to look at and how it will be measured. You then need to get as high a quality, comparable data from the parties as is possible before then undertaking the comparison.

Having compared performance (and for many people this is what they think benchmarking is – simply producing graphs and data), the next and in fact the most valuable aspect of a benchmarking exercise to undertake case studies to assess why the graphs look like they do. Difference in performance will be a function of two things:

- Environmental factors such as the age of the system, the funding streams, regulation and the characteristics of the urban area and,
- Manageable factors such as maintenance policy, operational policy, labour policy, investment policy, customer management etc.

What the exercise should do is to offer insights as to how influential the later are, and whether or not the methods can be transferred to improve performance. It must be said that if the environmental factors are by far and away more influential than the manageable factors there is little that management can do to improve performance. Change would be largely a function of external developments. Benchmarking exercises are therefore not suitable for everyone. The following hypothetical case studies illustrates the form of the exercise, and is indeed the third of the real life case studies described in a later section.

A graph may show that out of several publicly owned urban rail operators there is a noticeable difference in the reliability of the service and the levels of investment in fixed assets per annum. This would be expected. What would also be expected is that there would be a correlation between those that invest in their assets and service reliability. Yet, very often plotting the two against each other will not yield a clear correlation. At this point the analyst will ask why?

© Association for European Transport 2001

Well, to begin with some systems may be so old that the assets may simply be falling to pieces and there is little that can be done with investment short of patching them up. There may also be a mix up between investment and heavy maintenance – it is not unusual for the two to be confused and interchanged for the balance sheet. There may also be a problem with how reliability is measured and how comparable it is between metros.

However, aside from these there is the obvious question of where the money is being spent. There is also the issue of how much leverage new trains and signals could have over reliability if the operating discipline is poor and there is too much conflict between boarders and alighters and little adherence to scheduled platform dwell times. A benchmarking analyst will look at these issues - if they discover that the investment funds are not always project specific, but that in many cases the money is being spent upon visible projects such as new platform seats as opposed to critical aspects such as track and signals, and that those metros who don't do this have better reliability then the insight is useful. By probing the reliability issue further the analyst may also discover that those systems who take platform dwell seriously achieve levels of reliability over and above what their investment should deliver.

Such is the nature of the case study phase and to make this successful, the importance of solid research into the operating environment prior to the compilation of the performance graphs cannot be underestimated.

3.5 Implementation of best practice to improve performance

Having shown what some business and service providers do and how it affects their performance, some participants may decide to initiate pilot studies to see if they can replicate what has been done elsewhere. Thus one operator may decide to contract out some non-core services; another may decide to install count-down clocks for drivers on metro lines; another may move to a policy of preventative maintenance. Ultimately of course the decision to change has to be made by the management of the business and implemented throughout – other than advising on seeming best practice the activity of the benchmarking group is now complete (although of course the exercise should now begin again with a new focus – benchmarking is continuous). At this point it is worthwhile reiterating the fact that for the exercise to be effective it should be championed from the top – junior management will find selling, in some cases, seemingly radical methods from elsewhere difficult and getting the buy-in of all the organisation will be difficult.

Section 3 therefore has shown the steps involved in undertaking a benchmarking exercise and ultimately how the method can be used to implement change and improve performance. To illustrate this theory, the following sections will show practical example of where the technique has been used in transport across the world.

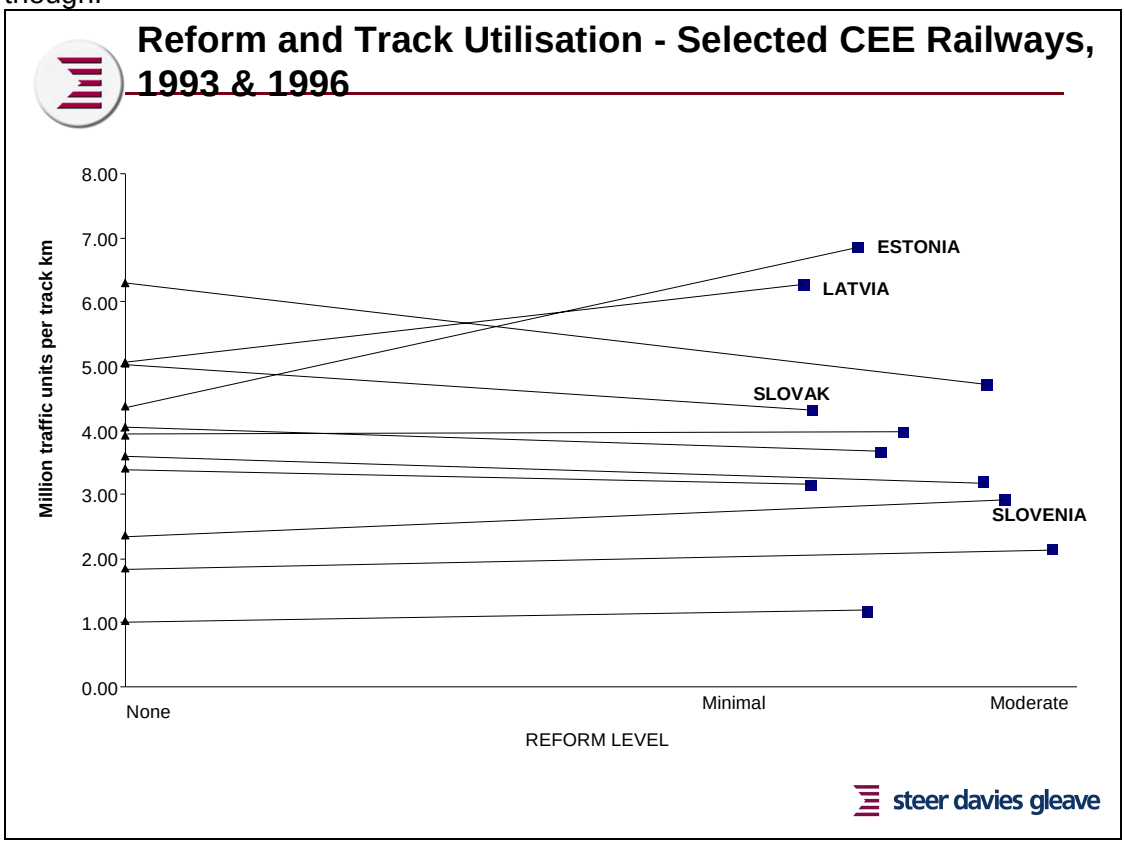
4 CASE STUDY 1 – EAST AND CENTRAL EUROPEAN RAILWAYS

In 1998 Steer Davies Gleave were commissioned by the Sumitomo Bank to produce a business plan for Slovakian Railways, ZSR, that would form a fundamental basis for a debt funding package. The railway was losing large sums of money, as its traditional markets declined after the collapse of communism and investment in higher quality road networks, and needed a strategy for turning things around. As part of the study the first named author of this paper carried out a benchmarking exercise of all the former East block railways using published data.

© Association for European Transport 2001

The hypotheses employed, and still employed to some degree, is that a deregulation of the railway in line with the provisos in EU directive 91/440 would allow the railway to improve performance. The benchmarking study therefore focussed upon the development of key performance indicators over time, whilst several of the railways were actively in the process of deregulation – separating accounts for operations and infrastructure, forming themselves into joint stock companies, reforming tariffs, reducing public service obligations, some even preparing for privatisation. Regulation was graded by means of a vector that marked the different attributes mentioned – this was a crude but useful measure. A complete summary of the exercise can be found in Ashmore and Mellor (1999).

A logical assumption to make would be that railways with greater commercial freedom can rationalise their supply network to meet demand in the most effective manner and therefore achieve higher levels of utilisation (and hence improve commercial performance). Utilisation can be measured in terms of traffic units per track km. Traffic units are the sum of the output measures, passenger kilometres and net freight tonne kms. Combining these output measures avoids distorting the analysis against railways who carry large volumes of freight, but not passengers and vice versa.
Figure 1 plots the reform grading against million traffic units per track kilometre for a selection of CEE Railways. The utilisation data relates to 1993 and 1996, 1996 being the latest comparable year for all concerned. The first noticeable aspect of Figure 1 is the vast majority of change, be it positive or negative has been moderate, and shows no correlation with reform. There are one or two exceptions though.



Latin Railways, LDZ's utilisation shows a marked increase, despite a small drop in traffic units, implying some network consolidation has taken place. The level of reform, however, has been moderate – there is still no contract for services provided under a public service obligation. One reason for the relatively high utilisation is the small size of the network. But the prosperity of LDZ between 1993 and 1996 was

© Association for European Transport 2001

primarily due to good links with Russia and aggressive marketing from the Latvian Ports. A positive spin off, for the Railway Company, of this relationship was that it was not forced to shed as many staff as other countries.
This has now changed due to a combination of factors – cooler relations with Moscow leading to the abolition of lower tariffs and increased competition from other Baltic ports. LDZ are in a much less promising situation in 1999 than they were in 1996. What does seem clear though is that the key driver at work in their prosperity is not reform.
The conclusion to be reached from this benchmarking exercise therefore is that in the case of Latvian Railways management actions will have less effect than political factors and a good relationship with Russia. Other comparisons show that a dense corridor network and a lack of alternative freight routes through other countries will also hold Eastern European Railways in good stead. Here benchmarking offered little in the way of optimism for performance improvement brought about by management, although of course sound investment in reliable infrastructure can never do any harm to a business.

5 CASE STUDY 2 – UK AIRPORT SURFACE ACCESS

In 1999 Steer Davies Gleave were commissioned by Paris CDG Airport to carry out a benchmarking exercise on dedicated rail links that serve airports so as to allow their own proposed link to learn from the successes and failures of others. A particular focus was on modal share of the dedicated links and a comparison of service attributes such as travel time, price and reliability. Much of the data again was taken from published sources such as the Civil Aviation Authority.

Figure 2 show the output from benchmarking mode share against distance from the city center. Intuitively, one would imagine that the further an airport is from the city centre the higher the access by car and taxi. In terms of the airports studied, this does not appear to be the case - if anything the relationship is inverse. Manchester airport, the closest to the city, centre has very high car access; Oslo, the furthest, has high access by rail. In Manchester's case this is almost certainly due to the number of package tourists from areas in the north of England other than Manchester who use the airport, couple with cheap taxis from the centre and it's excellent, free motorway links. In Oslo's case the fast dedicated rail link used by a high percentage of business travellers (who are less sensitive to cost than leisure travellers) offers faster, cheaper access than the expensive taxi service or the time consuming bus service.

© Association for European Transport 2001

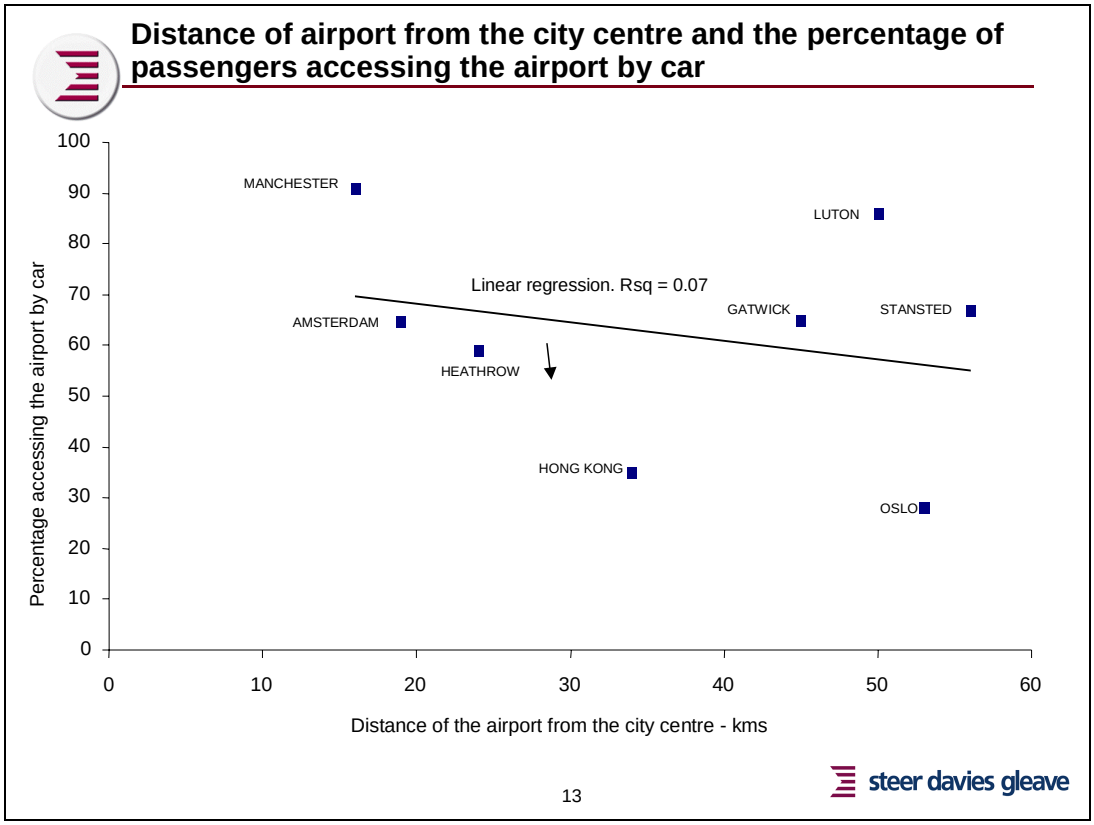


Figure 2 – Mode share versus distance of the airport from the city centre

When public transport mode share is benchmarked and plotted against price the effects are much clearer as shown in Figure 3. For those people using public transport, and therefore coming primarily from the city centre, the choice of either rail or bus is noticeably correlated with fare levels (as one would expect). Hong Kong, despite the high range of services offered by the new airport link, seems to have lower rail usage because the bus is so much cheaper - taxis would be very expensive. In the case of Amsterdam the bus and rail fares are almost identical, so the faster speed of rail probably drives modal choice.

© Association for European Transport 2001

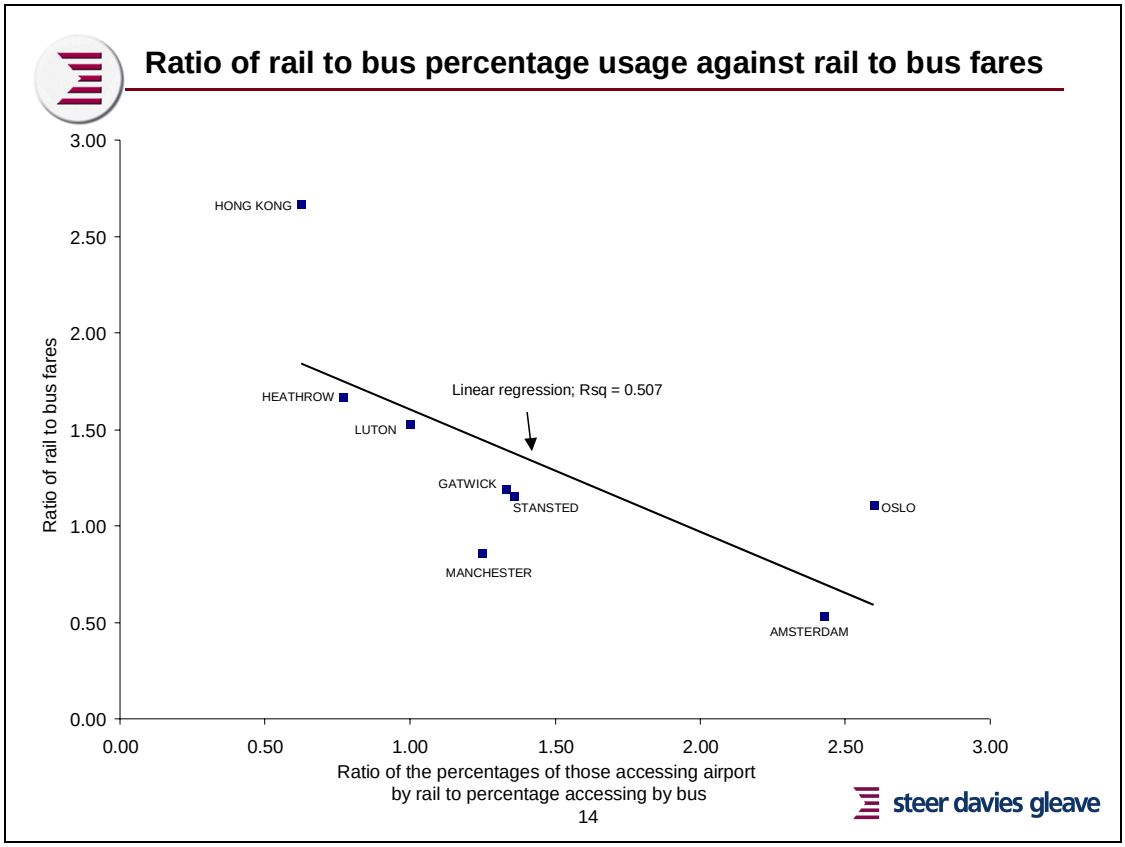


Figure 3 – Ratio of bus to rail fares against ratios of equivalent mode share

In conclusion therefore the benchmarking exercise showed that mode choice in airport transport does not appear to be dependent upon the distance from the city centre, but the type of trip, the access to the road network and the presence of a dedicated high-speed rail link. The public transport market seems most sensitive to price - this had the highest correlation coefficient with market share. Despite not being able to segment the data, it is intuitive that this effect is less significant for business travellers than leisure travellers. Journey time was shown to be less significant than expected. The market seemed to place greater weighting upon access to departure hubs and the frequency of the service involved. It is also probably fair to say that the market perceives rail as faster even though in some cases it may not be - due to the dedicated right of way.

6 CASE STUDY 3 – URBAN METRO SYSTEMS

The use of benchmarking is a well established management tool in many urban metro systems. At the forefront of the use of the technique were the management of the Mass Transit Railway Corporation of Hong Kong who have been using it for several years; this is shown in many of their annual reports.

Since 1995 the University of London Centre for Transport Studies has been running two benchmarking consortiums for several world metro systems, named CoMET and NOVA; for several years the first named author of this paper was an analyst on the study. For a full summary of their activities see Adeney and Self (2000). Many of the finding of the exercise have been used to directly influence management strategy and policy and have led to several pilot projects.

Figure 3 shows one of the key performance indicator graphs produced by the study. For reasons of client confidentiality there are no names attached to any of the

© Association for European Transport 2001

metros. What is apparent is the huge difference in service reliability between the Asian metros and the others. When one also goes onto examine the levels of investment that go into the systems it can also be seen that investment is not directly correlated with reliability. TO begin with some of the system reliability must be due to the newness of the assets – they simply don't fail as much. But even accounting for this the difference is staggering. When case studies were undertaken to probe a little deeper it became apparent that the Asian metros maintain their equipment differently, have rigid processes in place for system recovery to minimise downtimes and are fanatical in their management of platform dwell times.

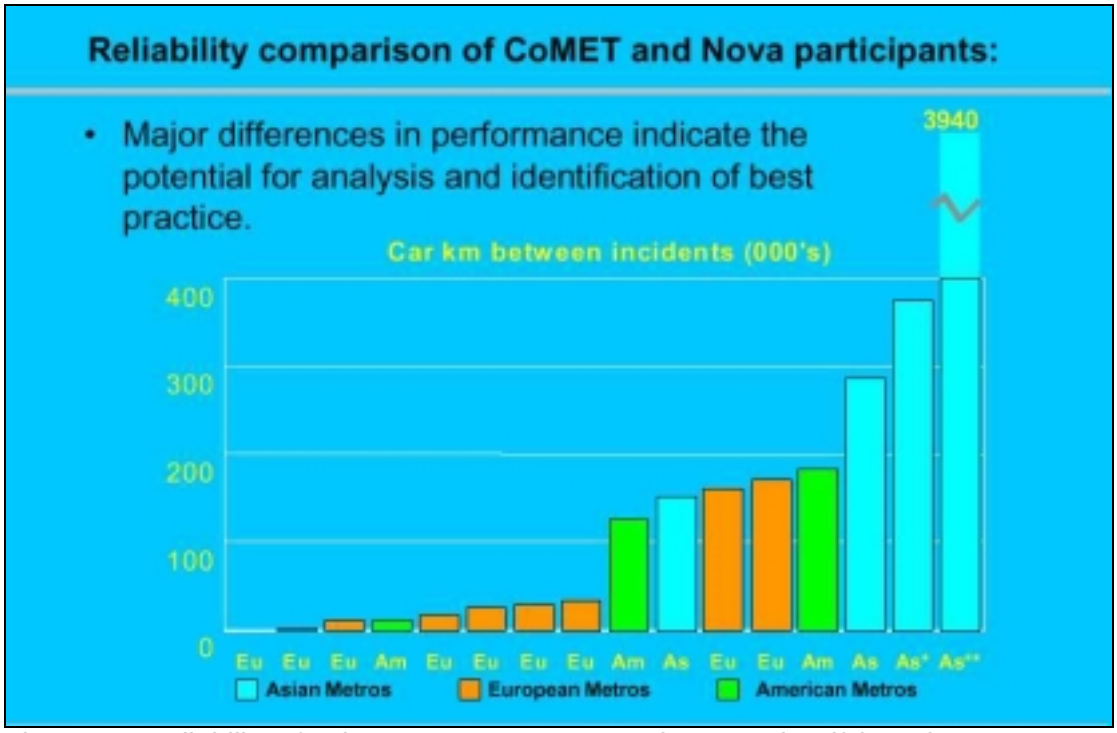


Figure 3 – Reliability of Urban Metros, Source – Adeney and Self (2000).

As a consequence of these findings an American metro instigated a platform queuing pilot project, with a marketing campaign and was able to increase the line capacity and reliability dramatically.

7 CONCLUSIONS

Benchmarking is a simple intuitive process, but it does not merely involve the production of loose comparison graphs.. It is a very powerful technique that uses the basics of common sense to identify why one similar business is performing better than another and then transferring the methods directly across to improve performance. It has been used successfully all over the world in the field of transportation and logistics. Often a small investment in a good benchmarking exercise can reap sizeable dividends – should anyone have any queries as to how benchmarking can help them then the authors would be delighted to hear from them!

8 REFERENCES

Ashmore D. P. and A. D. Mellor *To what extent can regulatory and structural reform improve the performance of east and central European railways?* Paper presented to the European Transport Forum, Robinson College Cambridge, September 1999.

© Association for European Transport 2001

Adeney, W and J. Self (2000), Benchmarking Metros of the World, Paper presented to the State in the Art of Benchmarking in the Transport Sector, proceedings located at http://bestransport.org/programme%20conference%202.htm.

© Association for European Transport 2001