Australasian Transport Research Forum 2012 Proceedings
28 - 30 September 2011, Perth, Australia
Publication website: http://www.patrec.org/atrf.aspx

An Assessment of Alternative Bus Reliability Indicators

By Currie G., Douglas N.J.[1] and Kearns I.[3]

Contact author Neil Douglas email: DouglasEconomics@ihug.co.nz
1 Graham Currie is Professor of Public Transport at Monash University
2 Neil Douglas is Manager of Douglas Economics Ltd.
3 Ian Kearns is Principal Manager, Performance Analysis at Transport for New South Wales

Abstract

Traditionally, there has been an emphasis amongst transport operators to collect statistics on operational performance such as % Buses Cancelled, % Departing On-time and % Arriving On-time. These 'operational' measures remain popular because the statistics are easy to understand and the data is relatively easy to collect especially if 'self-reported' by drivers. In addition, historical trends have usually been established and, as many operators use them, there is within industry comparability.

This paper has reviewed these operationally based statistics with seven other indicators of urban bus service reliability using an assessment framework and an assembly of evidence drawn from a study undertaken for the NSW Independent Transport Safety and Reliability Regulator (ITSRR) in 2006. The original 2006 review concluded that excess waiting time (EWT) was the preferred measure for high frequency bus services with the percentage of buses running on-time the preferred indicator for low frequency services with little to choose between measuring arrival versus departure times.

This paper has updated the 2006 review to take account of developments in the automatic collection of travel time data on buses and trains. For rail in Sydney, the incorporation of control systems that record the time of day at which trains are at stations has enabled a new measure of customer journey time delay (CJTD) to be developed that incorporates 'at station' and 'on train' delays and can be calculated for both frequent and infrequent services.

When considered alongside the other nine reliability indicators, the new CJTD measure ranked equal top with EWT. Both measures were considered easy to understand and have a high customer focus. Where they differed was in terms of fidelity/objectivity and cost efficiency. EWT is a partial measure since it excludes on-board delay whereas CJTD is a total measure incorporating 'at stop' and 'on bus' delay. EWT is also only appropriate for high frequency services whereas CJTD is applicable to low and high frequency services.

In terms of data collection, EWT is more feasible since it only requires information at unlinked bus stops whereas CJTD requires stop to stop bus times which are far more onerous to collect. However with automatic collection of travel times, CJTD should become a practically feasible measure for bus as has already been demonstrated for rail in Sydney.

Acknowledgements: Much of the material derives from a 2006 study that reviewed an observation survey of bus service reliability undertaken at bus stops in Sydney designed by the Independent Transport Safety and Reliability Regulator (ITSRR) of NSW. The authors would like to acknowledge the assistance of Robert Ramsay of ITSRR and now of Transport for New South Wales, Professor Luis Ferreira of Queensland University of Technology and Dr Nick Hounsell of Southampton University in undertaking the 2006 study.

Keywords: Reliability Measures, Average Lateness, Excess Waiting Time, Customer Journey Time Delay, Satisfaction Ratings, Customer Complaints.

1. Introduction

Surveys of customer opinion have consistently shown that timetable reliability is a critical factor in service quality. A 2009 survey of Sydney bus users by the Independent Transport Safety and Reliability Regulator (ITSRR) found 88% of respondents considered that *'buses keeping to timetable'* was important or very important (ITSRR 2009). In the UK, a national survey of rail passengers by MVA ranked service punctuality first out of thirty attributes in importance in 2005 and third in 2006 (MVA 2007). For Sydney, a 2006 survey found reliability to the dominant factor in explaining rail passengers' overall rating of service accounting for 25% of the overall rating (Douglas and Karpouzis 2006) and in the UK, the bus group First, found reliability to be even more dominant, explaining 34% of passengers' overall service quality rating (Balcombe et al. 2004). Thus timetable reliability is important and a main determinant of overall customer satisfaction.

For a bus and rail passengers, reliability covers two components: the reliability in arrival/departure time at the bus stop and the reliability in the travel time spent on the bus. To date, most reliability indicators have measured one of the two components but rarely both. Moreover there has been a tendency for objective reliability measures to focus on the bus rather than the passenger.

Despite the significance of reliability indicators only a few studies have examined the merits of alternative measures of reliability performance (Kittleson & Associates 2003; Mazloumi E et al. 2008; Trompet et al. 2011). None of these studies evaluated both objective and subjective indicators for both frequent and less frequent bus services. For example, Trompet et al. 2011 considered four objective measures for high frequency bus services. Moreover, none of the studies reviewed considered the value of indicators from the perspective of both a technical and non technical (i.e. bus passenger) audience.

This paper attempts to review a wide range of indicators of urban bus service reliability using an assessment framework and an assembly of evidence drawn from a study undertaken for the NSW Independent Transport Safety and Reliability Regulator (ITSRR) in 2006 (Douglas Economics 2006). This study reviewed an observation survey of bus service reliability undertaken at a sample of bus stops in Sydney by ITSRR which measured bus reliability in terms of bus cancellations and the percentage of buses departing on time. The 2006 review assessed the two measures against seven other indicators.

The 2006 review has been updated to take account of the developments in the automatic collection of travel time data by buses and trains. For rail in Sydney, the provision of automatic data has enabled a new measure of customer delay to be developed that incorporates at stop and on train delays. This paper has included this new measure into the review.

The paper is structured as follows; section 2 describes ten indicators of bus service reliability commenting on their information requirements, ease of understanding and measurement strengths and weaknesses. Section 3 then presents the evaluation framework developed to compare the measures and outlines the evaluation results. The paper concludes with a summary and assessment of the opportunities for further research in this field.

2. Reliability Indicators – A Review

The 2006 identified nine indicators that have been used as indicators of travel time reliability for bus services. The nine measures were:

- Percentage of Buses Cancelled
- Percentage of Services Departing 'On-Time'
- Percentage of Services Arriving 'On-Time'
- Excess wait time
- Average Lateness
- Service Variability Indicators
- Reliability Buffer Index (95 percentile on-board time / mean)
- Passenger Ratings of Reliability
- Customer complaints

The ITSRR surveys that reviewed calculated the percentage of buses cancelled and the percentage of buses departing 'on-time'. All but the last three reliability measures (passenger ratings and complaints) could be calculated using fieldworker observation surveys similar to those undertaken by ITSRR although the reliability buffer would require paired sets of bus stops in order to measure the variability in 'on bus' travel times.

Since the 2006 review was undertaken, ITSRR and latterly Transport for NSW (TNSW) has developed a new measure of customer delay for rail travel that has utilised the introduction of automatic vehicle location (AVL) devices on trains. These AVL devices provide actual station to station running times that can be used with patronage data to calculate a reliability measure that includes both 'at station' and 'on train' delays. This paper argues that with the introduction of similar data collection devices onto buses, a customer delay measure could be developed for Sydney bus services in the near future especially so it supplemented by electronic ticketing patronage data.

The remainder of this section describes each of the nine indicators reviewed in 2006 followed by the new customer journey time delay.

2.1 Percentage of Services Cancelled

For contracted services, cancelled services will generally have financial repercussions for the operator. It is therefore a statistic that is often mandated by tendering authorities.

For passengers, cancelled bus services can impose extra waiting time on passengers, particularly so on bus routes with infrequent services. On busy routes, cancelled services may increase crowding on following services, increase loading and unloading times, cause bus bunching and headway irregularity.

The percentage of buses cancelled is an easy to understand measure for the public, bus operators and planning authorities. However, the indicator may be difficult to estimate with accuracy by observation survey largely because cancelled services should be rare events. For example, Sydney Buses between August 2003 and July 2004 reported a cancellation percentage of 0.5% i.e. one bus was cancelled out of 200 scheduled services. Thus if a fieldworker 'missed' one bus in two hundred they observed, the cancellation rate would be double the actual figure. Thus the accuracy of fieldworker observations will depend on the diligence of fieldworkers. There may also be tendency to overestimate cancellations if late buses arrive after the end of the survey period.

There is also the definitional question as to what constitutes as a cancelled service. For example, buses that do not stop because they are full or running late could be counted as cancelled. For rail, these events are usually defined as 'skipped stops' rather than cancelled services.

There is also the possibility that buses that get 'out of series' may be mis-classified as cancelled rather than being early or late. In this regard, Transport for London (TfL) adopts an 'eight minutes early' and 'fifteen minutes late' as cut-offs; outside of which, buses are defined as 'not-linked' in their analysis rather than cancelled.

Working in an opposite direction, is the effect of duplicate or replacement bus services that may be sent out by operators to cope with heavy demand or services that were cancelled earlier.

The accuracy of observation surveys could be improved by having two or more observations per route; using video cameras so that observations can be double-checked and having a 'buffer' period of say 30 minutes at the end of the period to capture late running buses.

Consequently, given the difficulties of measuring bus cancellations by observation survey, such statistics usually rely on self-reporting by bus operators who usually measure cancellations in terms of the number of buses leaving or arriving back at the depot using reports compiled by drivers.

2.2 Percentage of Services Departing On-Time

The percentage of buses departing on-time is a widely used indicator of punctuality (Strathman et al. 2000). It is an indicator that is easy for both management and the public to understand.

As well as measuring timetable adherence, the percentage of buses departing on time may be viewed as an indication of bus stop waiting times: higher percentages indicating shorter waiting times; lower percentages - longer waiting times. In this regard, the indicator is more likely to be relevant for infrequent services when passengers time their arrival at the bus stop rather than frequent services when passengers turn up at random.

There is therefore the question of how infrequent does a bus service have to be for bus punctuality to be important. A benchmarking study by Trompet et at (Trompet et al. 2011) of regularity indices drew a threshold of ten minutes between frequent and infrequent services: *"although a universal definition for the headway length to mark the difference between high and low frequency routes could not be found in the literature, a 10-min headway seems the most-used boundary"*.

To calculate the indicator, 'linkage rules' are required that link the actual time of departure to the scheduled time. For scheduled time, ITSRR uses scheduled times published on the web (Douglas Economics 2006). A key decision is what 'on-time window' should be adopted for services to qualify as departing 'on-time'. ITSRR adopt an on time window of no more than 2 minutes early and 5 minutes late. This window was the same as that adopted by the Department of Infrastructure (DoI) in Victoria and TfL for low frequency services (Department of Transport 2008; Transport for London 2012).

A pilot survey undertaken in October 2009 by the NSW Transport and Infrastructure (NSWTI) adopted a tighter one minute rule for early departures (Achterstraat 2010).. The NSWTI rule was that services should leave no more than one minute early and no more than five minutes late.

A survey of service delivery measures collected by transit agencies by the US Transportation Research Board (TRB) showed zero minutes to be the most common earliness threshold and five minutes to be the most common lateness threshold (Kittleson & Associates 2003). Commenting on the zero early threshold, this report stated "*vehicles departing before their scheduled departure time are almost never considered on-time"*. An earlier but larger survey of 83 agencies in the USA found that most agencies adopted a threshold of 1 minute early and 5 minutes late as ‘on-time’ (Howard 1995).

Contrastingly, Sydney Buses treat buses running early as ‘on-time’ since on-time is measured at the terminus when early arrivals should benefit or, at least not disadvantage, passengers (Sydney Buses 2006).

For trains, Melbourne Australia reports the percentage of services that arrive on time at destination with trains deemed on time if they are between 59 seconds early and four minutes and 59 seconds late of the scheduled time. Financial compensation (in the form of free tickets) may be paid to eligible customers if less than 88% of trains are on time (Department of Transport 2008).

In theory it should be possible to report the percentage of buses running on-time for different thresholds. Rail operators often report the percentage of trains running on-time for two thresholds. RailCorp NSW reports the percentage running up to 5 minutes late and up to 10 minutes late for suburban CityRail services. Queensland Rail logs trains running up to 3 minutes late and up to 5 minutes late.

Sydney Ferries reports an ‘on-time running rate’ which measures the proportion of actual services departing within 5 minutes of the scheduled departure time from the first wharf on the service (Sydney Ferries 2012).

For infrequent services, separating early from late running buses has merit since substantial early departure times are particularly bad for passengers who time their arrival at the bus stop only to find that the bus has already departed.

Bus stop location is likely to influence on-time running. ‘Start’ bus stops are more likely to run to time than mid-point bus stops. The NSWTI pilot survey recorded the time buses departed major interchanges, which *“can be misleading, as it is not taken when buses are en-route” (Achterstraat 2010)*.

In terms of calculation, the percentage of buses departing on-time can be expressed in terms of either the percentage of buses scheduled or the number of services actually operated; the difference being the number of cancelled services (allowing for any duplicated services). However, given that the number of cancelled services is usually relatively very small (typically 1% of scheduled services), the difference in the two measures should rarely be of consequence. In the ITSRR survey, the percentage was calculated in terms of the number of buses operated.

For high frequency services, measuring departure time punctuality is problematic especially so for services that are ‘timetabled’ in terms of regularity such as *“ buses every 5 minute between 7am and 9am”* but with specific departure times not given. Passengers are then concerned with waiting times being no greater than that indicated by the timetable rather than individual bus punctuality.

Instead of measuring punctuality for individual bus services, the reliability measure for high frequency bus services is be better specified in terms of headway (the interval between successive services). The measure would then be the percentage of headways within a specified range of that scheduled. Thus, if the timetable specifies a headway of 5 minutes (i.e. ‘buses every 5 minutes’), the measure might be the percentage of headways within X minutes and greater than Y minutes of 5 minutes. For example, Barcelona MTB calculates the percentage based on 1 minute less and 3 minutes more than the scheduled headway and Vancouver CMBC uses 2 minutes less and 4 minutes more (Trompet et al. 2011). An alternative to calculating the measure in terms of an absolute deviation is the percentage deviation such as ±20% of the scheduled headway.

2.3 Percentage of Services Arriving On-Time

The alternative to measuring punctuality in terms of departure time is to use the arrival time at the stop. Sydney Buses measures punctuality in terms of the arrival time at the terminus.

Often, there will be little difference between the departure and arrival times at bus stops, especially if times are recorded to the nearest minute. There are some advantages and disadvantages to each approach as shown in Table 1.

The percentage of buses running early can only be measured in terms of departing times. Buses arriving early represent a bonus for alighting passengers, but need not be a bonus for passengers continuing on the bus if the bus waits until the scheduled departure time (implying too much slack in the timetable). Buses departing early however can be a significant cost to joining passengers since some may miss the bus and have to wait for the next one.

At designated timing points and driver layover points, measuring bus departure times is preferable since it is more consistent with operating schedules. At timing points, measuring arrival times would make on-time running look worse than measuring departure times. Furthermore, there may be no scheduled arrival time as such thus measuring actual arrival times against scheduled departure times will result in biased measures with too high a percentage of 'early' buses.

Obviously, there can be no arrival time for the bus stop at the start of a route and no departure time at the end of the route.

Table 1: Relative Advantages of Measuring Arrival & Departure Times

Criteria	Best Measure	Comment
1. Percent Early Running	Departure	Arrival times are not appropriate for measuring "early" running buses
2. Operator Schedules	Departure	Departure times are more consistent measures at layover/timing points where operators plan services
3. Route Start Stop	Departure	Departure times are appropriate for the start stop
4. Passenger Focus	Arrival	Customers are probably generally more focussed on bus arrival times
5. Measurement Accuracy	Arrival	Arrival times are easier to record accurately as they are not affected by passenger boardings
6. Route End Stop	Arrival	Arrival times are appropriate for the end stop

Recording arrival times should be easier at bus stops that have no 'catch-up time' allowance. Departure times will be complicated by late arriving passengers rushing to the bus requiring the bus driver to re-open the doors etc. Trompet et al found that actual arrival times were collected more often than departure times by bus operators who are members of the

International Bus Benchmarking Group (IBBG) thus facilitating performance comparison (Trompet et al. 2011).

Likewise, rail operators tend to focus on arrival times. RailCorp NSW measures on-time running in terms of the final destination station which in the AM peak is the main CBD station. Queensland Rail also measures on-time for running in terms of a single terminus station.

From a passenger perspective, measuring arrival times has advantages. Bus passengers are more likely to stop measuring their waiting time when the bus arrives (rather than until the bus departs even if it sits at the stop for several minutes). Most behavioural research has shown that passengers attach a lower cost to the time spent onboard the bus than waiting for the bus.

One study however suggested that departure times should be preferred where heavy ridership causes lengthy dwell times for passengers waiting to board (Nakanishi 1997).

Theoretically, it should be possible to record both the arrival and departure time but it is not always obvious which time should be compared with the timetable. There can also be a practical problem in recording both times if buses set down passengers at one stop then move to another stop to pick up passengers and then depart.

2.4 Excess Wait Time

Excess wait time (EWT) results when the interval between buses is longer than indicated the timetable indicates; it has been used in London for high frequency bus routes that operate on a headway basis. Transport for London (TfL) defines bus services as high frequency when there are five or more buses per hour, i.e. operating to headways of no more than 12 minutes. Low frequency services are defined as four or fewer buses an hour (Transport for London 2012).

London Buses has estimated EWT by observation survey undertaken at around 500 locations across London. Each location is surveyed 16 times over a 12 week period with each survey lasting for up to three hours. This equates to around 40,000 hours of survey data per year which in 2006 required a team of over 100 surveyors.

The underlying assumption for EWT is that for such high frequency services, passengers tend not to look at the timetable but turn up at the bus stop at ‘random’. For this reason TfL considers that customers are “more interested in how reliable and evenly spaced the service is rather than the punctuality of specific buses: *“we aim to ensure buses run at evenly spaced intervals and do not bunch.”* With high frequency bus services, the scheduled waiting time (SWT) can be shown to be half the headway between buses. Thus with a headway of 12 minutes, the SWT would be 6 minutes.

In practice, the actual waiting time (AWT) at a particular bus stop will differ from the scheduled waiting time (SWT). The AWT will depend on the frequency and the regularity of actual bus arrivals at the stop in question. The AWT can be estimated by undertaking surveys of the actual arrival time of buses at stops.

EWT can then be calculated as the difference between the AWT and the SWT, that is EWT=AWT-SWT.

The ratio of AWT over SWT can be calculated to indicate how much longer, on average, passengers are waiting than intended by the scheduled headway. A ratio of 1.5 for example would indicate that passengers waiting 50% longer than they the schedule would indicate.

The aim behind publishing EWT figures is to provide bus passengers with the information to plan their journeys better. Accordingly, TfL publishes EXT wait time figures for all high frequency bus routes in London on a quarterly basis. As well as reporting EWT, TfL also presents the percentage chance of waiting less than ten minutes, 10-20 minutes, 20-30 minutes and more than 30 minutes.

It should be mentioned here that EWT is only a partial measure since it does not include on-bus travel time delays. EWT is also a less useful measure for less frequent bus services where passengers tend to time their arrival at the bus stop to meet a particular bus service. For London, two-thirds of bus services were defined as high frequency and one third as low frequency in 2006. For these low frequency services, indicators that measure adherence to the timetable are more useful. For London, the percentage departing on time is measured with buses classified as early if they depart between 2 minutes and 8 minutes before the advertised time. Of course, some buses in this category may be late running buses, which could be regarded by passengers as the next bus running early. Buses are classified as late if they depart between 5 and 15 minutes after the advertised time

2.5 Average Lateness

Average lateness incorporates the degree of lateness into the reliability measure and can be calculated as the percentage of buses late multiplied by the number of minutes late. For example, if 20% of buses were 5 minutes late, average lateness would be one minute (0.2 x 5). If 15% of buses were 5 minutes late and 5% 10 minutes late, average lateness would be 1.25 minutes (0.15 x 5 + 0.05 x 10). Thus, the measure is less crude than the percentage of buses arriving / departing within an ‘on time window’ though there are concerns that by presenting only an average measure, the variability of travel times and hence the lack of travel time predictability is lost.

Average Lateness should be able to be calculated with no more effort than the percentage of buses on time. However, the measure is only strictly appropriate for services that run to a timetable. For frequent bus services that run to a headway, the measure would be difficult to calculate other than by computing excess waiting time. In fact, other than in the guise of excess waiting time, average lateness has been rarely reported by transport operators.

The measure is sometimes used in rail demand forecasts to translate changes in reliability into equivalent on-board travel time, see for example Balcombe (2004). The measure has particular merit if it is calculated for the whole trip thereby including waiting plus on-board time lateness (see section 2.10).

It is worth mentioning, that for timetable services, median lateness might be a better estimate than the mean because of the distorting impact of outlier observations. This choice is supported by a review by Quiroga who concluded that for travel times, the median tends to be favoured over the mean (Quiroga 1997).

2.6 Statistical Measures of Variability

The variability in actual arrival, departure or on-board bus times can be summarised using statistical measures of variability such as variance, standard deviation and the coefficient of variation. Unfortunately, such measures are unlikely to be well understood by bus users and or bus operators.

For frequent bus services (headways of 10 minutes or less), the *"US Transit Capacity and Quality of Service Manual"* advocates headway adherence as a preferred measure of service reliability (Kittelson & Associates et al. 2003). The measure is calculated as the coefficient of variation of bus headway:

$$c_{vh} = \frac{\sigma}{\text{mean scheduled headway}} \quad .....(1)$$

where: c_{vh} = coefficient of variation of headways and

$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(AHway_i - SHway_i)^2}$ = standard deviation of actual – scheduled headways

Headway deviations are measured as the actual headway minus the scheduled headway. The resultant coefficient of variation can then be expressed as a probability that a given transit vehicle's headway will be out by more than half the scheduled headway. Table 2 taken from the US Manual correlates level of service (LOS) with headway adherence (c_{vh}). A low coefficient of variation of bus headway of under 0.21 would indicate that the service is provided like clockwork. A figure of 0.4-0.52 would indicate irregular headways with some bunching and a value of 0.75 or greater would indicate most buses are bunched.

Table 2: Headway Adherence

LOS	c_{vh}	$P(h_i > 0.5\,h)$	Comments
A	0.00-0.21	≤1%	Service provided like clockwork
B	0.22-0.30	≤10%	Vehicles slightly off headway
C	0.31-0.39	≤20%	Vehicles often off headway
D	0.40-0.52	≤33%	Irregular headways, with some bunching
E	0.53-0.74	≤50%	Frequent bunching
F	≥0.75	>50%	Most vehicles bunched

NOTE: Applies to routes with headways of 10 minutes or less.

Source: US Transit Capacity and Quality of Service Manual – Exhibit 3-30

A simpler measure referencing only the standard deviation of the difference of scheduled and actual headways (the numerator in equation 1) has been used by Trompet in a benchmarking study of higher frequency bus services (Trompet et al. 2011). Trompet saw the lack of normalization by dividing by the mean scheduled headway as a drawback but one offset by greater simplicity and the the fact that the resultant measure is expressed in minutes. Thus, a calculated standard deviation of 2 minutes means that roughly two thirds (68%) of headways would be within two minutes of the scheduled headway (assuming a normal distribution).

The UK Passenger Demand Forecasting Handbook (PDFH) for rail services uses standard deviation of lateness in a measure called the reliability ratio that divides the standard deviation by average lateness (Association of Train Operating Companies 2002).

2.7 Reliability Buffer Index

The reliability buffer has been used for road congestion monitoring and reporting by the US Department of Transport Federal Highway Administration (FHA 2004). The reliability buffer expresses the 95th percentile 'commute' time in relation to the average 'commute' time, equation 2.

$$\text{Reliability Buffer Index} = \frac{\text{95th Percentile Time - Average Time}}{\text{Average Time}} \quad .....(2)$$

The idea of the index is that bus passengers will build in a 'buffer' into their trip planning to account for the variability. By building in a buffer, bus passengers will arrive early on some days. This may not necessarily be a bad thing, but the extra time is still carved out of their day - time they could be using for other pursuits besides commuting."

Of course, most bus operators will build in a reliability buffer into their timetable by adding a time allowance at one or more timing points. Thus the observed measure will be affected by how much slack operators build into their timetables.

The buffer can be used to measure travel time reliability. Several statistics can be developed from the information. The FHA considered the Buffer Index to be a particularly useful one. The buffer index is calculated as the extra travel time needed to accomplish a trip 19 times out of 20 chances in relation to the average travel time for that trip. Tracking changes in the Buffer Index over time indicates whether reliability is improving or deteriorating.

The information requirements for the reliability buffer are similar to those for average lateness but the measure is less easy to understand. Comprehending what '1 in 20 buses taking 10% longer than the average bus' means is more convoluted than 'buses are 2 minutes late on average'. On the other hand, the reliability buffer does try to get across the message of the variability in travel time whereas average lateness does not.

There is also the question of why the 95th percentile? Possibly, it is used because of the convention of using the 95% confidence level in statistical tests. For passengers it would mean being late to work or home once every two weeks. As an alternative to the 95th percentile, Bullock et al used the 85th percentile in a trial of GPS data loggers in Sydney (Bullock et al. 2005).

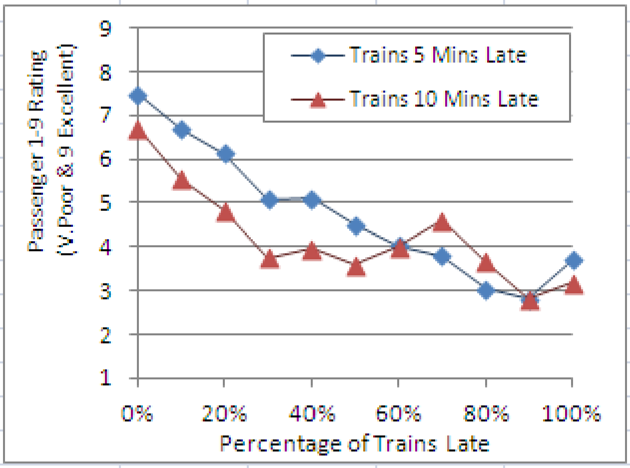
2.8 Passenger Satisfaction Surveys

An alternative to measuring actual bus service reliability is to survey public opinion. Indeed, in a review of rail service reliability measures, ITSRR considered that customer satisfaction sourced from customer surveys was the most frequently used customer measure of delay (ITSRR 2008). Accordingly, there are many examples of customer satisfaction surveys of which three Sydney surveys are reviewed below.

A 2005 survey combined experience of late trains with a satisfaction rating question (Douglas Economics 2006). The survey involved a self-completion questionnaire survey of over 1,000 Sydney rail passengers surveyed between October 2005 and January 2006. Passengers were asked how many rail trips out of the last ten were over 5 minutes late and over 10 minutes late, Respondents were then asked to rate the reliability of their rail service on a 1-9 scale with 1 being very poor and 9 excellent. As might be expected, passengers who experienced a high percentage of late trains rated reliability low and passengers who experienced few late trains had a higher rating. Figure 1 shows the relationship. In fact, the relationship might be considered rather insensitive since the highest average rating only reached 7.5 for passengers who did not experience a single train over 5 minutes late and fell to only 3.8 for passengers who reported that all their last trains had been over 5 minutes late. Such a lack of sensitivity points to problems in communicating actual performance to passengers through subjective rating measures.

In 2009, ITSRR undertook a telephone market research of just over 2,300 bus users across the metropolitan region of Sydney, ITSRR (2009). The survey followed on from similar annual surveys of rail user satisfaction conducted between 2004 and 2009 (ITSRR 2010). One of the attributes was 'buses keeping to the timetable'; the other attributes included cleanliness, crowding, bus connections, timetable information at stops and driver friendliness. Obviously, the ability of surveys to cover a range of service attributes is a distinct advantage.

Figure 1: Satisfaction Rating & Experience of Reliability
Survey of 1,000 Passengers 2004-5



Bus users were asked to rate the importance and quality of ten aspects of service on a five point scale ranging from 'not at all important' to 'very important' on the importance scale, and from 'very poor' to 'very good' on the quality scale. The response to the two questions provided up to 25 response combinations (5x5). Satisfied bus users were required to rate the attribute as 'acceptable', 'good' or 'very good' and also rate the importance of the attribute as 'desirable', 'important' or 'very important'. For 'buses keeping to timetable' 77% of respondents were 'satisfied'. 'Dissatisfied' bus users were those who rated the attribute 'poor' or 'very poor' and rated the importance as 'important' or 'very important'. 19% of respondents were dissatisfied with 'buses keeping to timetable'. The remaining 4% were respondents who considered 'buses keeping to the timetable' as not at all important or somewhat important. Thus the importance adjustment was not marked for bus reliability.

In 2011, the Bureau of Transport Statistics (BTS) undertook a customer survey of train, bus and ferry customers using a standardised on-board self-completion survey (BTS 2011). The survey achieved a response rate of 60% with 12,323 train, 6,325 bus and 3,946 ferry completed questionnaires returned. Passengers were asked to rate 23 individual attributes of service and give an overall trip rating. Services being on time (keeping to timetable) was one of the attributes. Satisfaction was measured on a five point scale of 'very dissatisfied', 'dissatisfied', 'neither', 'satisfied' and 'very satisfied'. The survey found 27% of bus users were dissatisfied; 18% neither satisfied nor dissatisfied; and 55% satisfied with bus services being on time. Thus, over the two years 2009-2011, satisfaction with buses keeping to timetable fell noticeably from 77% measured by the ITSRR telephone survey to 55% measured by the BTS self-completion survey. How much of the fall was due to actual operational performance versus change in survey method (from a telephone interviewer survey to a self-completion on-board train survey) has not been assessed.

Although, satisfaction surveys are necessarily subjective, the data can provide a useful check on operational data provided by operators. In a review of public transport service reliability, ITSRR found that the results of its 2009 bus customer satisfaction survey "were not always consistent with operational performance data" and that "such inconsistencies add to ITSRR's concerns regarding the quality of the operational performance data reported by various operators to the Ministry of Transport" (ITSRR 2010).

2.9 Customer Complaints

Passenger complaints which mostly results from personal experiences such as *"my bus has been late three days in a row" are* a useful indicator of service reliability.

The public should relate well to complaint statistics and 'indices' such as the number of complaints per thousand scheduled bus services or per thousand bus trips. The disadvantage is that complaints are not an accurate measure of actual on-time running. Instead they reflect 'annoyance', the willingness of the public to report problems and the ease by which problems can be reported. Usually all complaints are given an equal weight irrespective of their severity (although weights could be applied to different types of complaint).

As with customer satisfaction surveys, it should be possible to correlate complaints with actual on-time reliability statistics thereby providing a check on operator statistics.

Complaints can be categorised under headings. Between July 2005 and June 2006, the NSW Ministry of Transport categorised feedback into 19 'complaint' categories and one compliment category in a submission to the Independent Pricing and Regulatory Tribunal (Ministry of Transport NSW 2006). Of the complaint categories, four were reliability related: bus early, bus late, missed stop and running wrong route. Of these categories, bus late and missed stop were the most frequent complaints averaging 587 (53%) and 347 (32%) per month out of a total of 1,101 reliability related complaints. Early buses 120 (11%) and running the wrong route 47 (4%) were less frequent complaints. Total complaints averaged 2,166 per month. Therefore, reliability related complaints were around one-half of all complaints.

Where a centralised complaint logging system is set up, the data can be collected and processed cost effectively. It should also be possible to analyse the number of complaints by route, operator or time period and develop indices that relate the number of complaints to the number of scheduled bus services and/or passengers.

2.10 Customer Journey Time Delay

Customer journey time delay measures the difference between the customers expected and actual travel time from the start to the finish of the bus trip. Unlike the previous objective indicators, the time spent waiting for the bus as well as travelling on the bus is included so that the measure represents the travel time reliability of the whole bus trip. To do this a customer's bus stops (both origin and destination) need to be known as well as the timetabled services and actual services operating. Moreover, departure and arrival times for the origin and destination bus stops need to be combined, both scheduled and actual.

Calculation of the measure is therefore necessarily far more complex than measuring arrival or departure time reliability since stop-to-stop running times are required. Either pairs of bus stops need to be surveyed perhaps using number plate matching or onboard travel time surveys undertaken. Either of these manual method would be resource intensive.

However, the increasing use of automatic vehicle location devices on buses should provide stop-stop running times at near zero cost. There is also the possibility of using data from smart card ticketing which can record the time passengers 'tag on' and 'tag off' at stops. This ticketing data would be particularly useful since it provides a way to calculate customer rather than vehicle delay although there remains the issue of calculating passenger wait times for infrequent services.

For rail, a measure of customer journey time delay has been pioneered by ITSRR for Sydney rail services (ITSRR 2004). Late arriving services are included in the measure and the impact of early departing services is included. The measure also accounts for trains that 'skip stops', that is, miss a timetabled stop. To calculate customer delay for a specific passenger, two assumptions need to be made. The first assumption is that the passenger is aiming to use the first train service scheduled to arrive at the origin station after the passenger's arrival (and which stops at the passenger's destination station). This train is referred to as the 'scheduled train'. The second assumption is that the passenger actually catches the first available train, that is, the first train to arrive at the origin station after the passenger's arrival (and which also stops at the passenger's destination station). This train is referred to as the 'actual train'. Customer delay is the difference between the actual arrival time at the destination station of the actual train and the scheduled arrival time of the scheduled train at the destination station. The actual train and the scheduled train may be one and the same train service or may be different train services.

In Sydney, information on train arrival times is provided automatically by a Train Location System (TLS) which records the scheduled and actual times of trains as they reach certain locations on the rail network. Figure 2 illustrates the calculation of customer delay with some hypothetical data. The table lists eleven trains connecting Bankstown and Central in the first column. The alternating lighter and darker bands of colour indicate the scheduled and actual trains for six passengers.

Figure 2: Example of Customer Delay

		Scheduled train		Actual train			
	Passenger arrival time at Bankstown	Bankstown departure time	Central arrival time	Bankstown departure time	Central arrival time	Customer delay (minutes)	Train delay (minutes)
1	7:19	7:20	7:47	7:23	7:49	2	2
2	7:29	7:29	8:02	7:28	8:04		2
3		7:39	8:11	7:39	8:15	13	4
4		7:50	8:17	7:55	8:18		1
5	7:59	8:00	8:32	8:03	8:26	0	0
6		8:09	8:41	8:11	8:44	0	3
7	8:10	8:20	8:48	8:34	9:02		14
8		8:38	9:10	8:40	9:12		2
9	8:45	8:46	9:18	cancelled service			–
10		9:00	9:32	9:04	9:35	17	3
11	9:10	9:12	9:35	9:12	9:35	0	0

Note that, in this example, there are no passengers that either aim to or actually catch trains 4 and 8.

Passenger 1 arrives at the station at 7:19 to catch the scheduled 7:20 train. The customer catches that train but it actually departs Bankstown at 7:23 and arrives at Central two minutes later than scheduled. Both the train and the customer are delayed by two minutes.

Passenger 2 arrives at the station at 7:29 to catch the scheduled 7:29 train but misses it because it departs one minute early at 7:28. The customer catches the scheduled 7:39 train which departs on time from Bankstown but arrives at Central at 8:15, four minutes later than the scheduled time of 8:11.The customer is delayed by 13 minutes compared to the scheduled arrival time (8:02) of the train the customer was aiming to catch.

To calculate customer delay across the rail network, estimates of the number of passengers arriving per minute at the origin stations are required. These can be approximated by barrier counts. An origin-destination trip matrix is also required to give the number of passengers travelling between stations.

As mentioned above, a customer journey time delay measure could be calculated for buses but would require similar detailed information covering bus travel times, passenger arrivals at bus stops and information on where passengers are travelling in order to weight the index. In this regard a Singapore study is of relevance which has used data from the electronic ticketing system to give actual bus arrival times at bus stops and thereby monitor bus headway and punctuality (Mezghani 2008).

3. Evaluation Framework

3.1 Previous Evaluations

Despite the obvious importance of bus service reliability, there have been few comparative assessments of public transport reliability indicators and those that have been conducted have been limited in scope.

A recent benchmarking study of regularity measures focussed on higher frequency bus services in major cities and evaluated four measures (excess wait time, standard deviation of the difference between the actual and scheduled headways, wait assessment +/- 2 mins of scheduled headway and service regularity +/- 20% of scheduled headway). The evaluation framework used five comparison criteria: ease of communication; whether the measure was subjective or objective; customer representation, condition requirements of the measure and whether the measure penalizes very long headways (Trompet et al, op cit). The research found that whereas all methods illustrate a different interesting view of regularity performance, excess wait time was preferred since it reflects customer experience of service regularity. While the study is of much value, it only considered a limited range of indicators for higher frequency bus services.

An assessment of the merits of alternative measures of tram reliability performance was also undertaken in Melbourne (Mazloumi E et al. 2008) however in this case the research focussed on measures of travel time variability and the application of Automatic Vehicle Monitoring (AVM) database records in assessing the performance of measures. Measures of Travel Time Reliability (TTR), that concern the variation of TTs in relation to an expected value, and Travel Time Variability (TTV), the degree of variation demonstrated by AVM data were evaluated against a large data set of AVM records. The focus of this paper was on technical measures of variability/reliability rather than measures that might be of use as a means to communicate with the wider public. Hence a technical focus was present in the analysis. The study found that, in general, there was no significant difference in TTV/TTR outcomes between the measures tested. Where variations emerged, it was due to the ‘sensitivity’ of measures to outliers in the data series. The study recommended The Buffer Index and the Coefficient of Variation to measure TTR and TTV, respectively, mainly on the basis of the approach to measurement, which provides additional information to users and planners. However this research found that the selection of time interval length over which TT observations were aggregated for analysis of TTR/TTV was critical to outcomes. Shorter interval selection was recommended. While this paper was of value, its focus was on measures of variability rather than measures that can be used to communicate reliability performance to the public. This is why the research presented in this paper has considered a wider range of performance measures and has used a more comprehensive evaluation framework including user perspectives.

3.2 Evaluation Framework

A framework has been used to assess the reliability indicators based on four criteria: (1) ease of understanding; (2) the extent to which the measure has a customer focus; (3) the accuracy, completeness and objectivity of the measure; and (4) the relative cost/effort in collecting and analysing the data.

Each of the four criteria is described below.

The purpose of a reliability indicator should be to provide an easily understandable measure (1) of how well buses and trains operate to a timetable so that the public can determine for themselves how much time they should allow to make certain trips in addition to that indicated by the timetable. Really, ease of understanding can only be assessed by the public and not by 'experts' on their behalf but in the absence of any consumer market research we have assessed the measures in terms of their simplicity.

Customer focus (2) recognises the need for transport agencies to report indicators that measure the 'customer experience' rather than that of the bus or train. Evidence for this need is recent research by the Metropolitan Transportation Authority in New York (MTA) which concluded that *'a true passenger based on-time metric still eludes the MTA and the other major U.S. transit agencies, except for BART (Bay Area Rapid Transit)'* (PCAC 2011). The MTA recommended that, as a priority, research be commissioned to develop such a metric. There is also the research of Frumin who has developed new metrics using information collected by the Oyster card in London so as to *gain 'a deeper and more nuanced appreciation of the passenger experience than is currently provided*' (Frumin 2008).

The third criterion considers the completeness, fidelity and objectivity of the reliability measure. As was mentioned in the introduction, service unreliability comprises delays at the bus stop and delays on the bus. Thus for a completeness, both components of delay need to be included. The measure should also offer a high degree of fidelity, i.e. provide an exact and non distorting measure of reliability as experienced by the passenger. The measure should also be objectively based so that comparisons between bus routes and operators can be made, trends in reliability established and the impact of timetable changes, infrastructure and policy identified.

The fourth criterion is the cost/effort of collecting, analysing and reporting the data. Since the 2006 review, the introduction of automatic travel time collection devices on trains has presented the opportunity to calculate reliability indicators for rail journeys in a comprehensive, up-to-date and cost effective way. With a similar roll out of automatic vehicle location devices for Sydney buses in the near future, a similar potential for collecting bus service reliability data will become available. However, the scale of reliability monitoring task for bus will be far bigger than for rail. As a indication of the scale of the problem, the aim of the 2004 and 2005 ITSRR surveys was to cover the 15 bus contract regions of Metropolitan Sydney that together cover around 5,400 square kilometres extending 90 kilometres north-south and 60 kilometres east–west in which 548 individual bus routes operate and connect a total of 23,000 bus stops. By comparison there are 307 rail stations operating on twelve rail corridors. Therefore some form of sampling of bus services to keep the monitoring exercise manageable will be required.

3.3. Evaluation Results

Table 2 presents the criteria and overall rating scores for the assessment of the measures examined. To calculate the scores, a three point rating scale was used in which a low rating was given a zero score, a medium rating one point and a high rating two points. An overall rating was then calculated by adding the scores for the four criteria.

In the original 2006 review, excess waiting time (EWT) and the percentage of buses running on-time were the preferred indicators with little to choose between measuring arrival or departure times. The ranking of the indicators was affected by the frequency of the bus service. For example, on high frequency routes, EWT is easier to calculate than the percentage of buses arriving or departing on-time due to difficulties in linking buses to the timetable.

In Table 2, the proposed new measure of customer journey time delay (CJTD) is ranked alongside excess wait time (EWT) is as the best overall indicator. Both measures achieved a high overall rating with; both considered easy to understand and having a high customer focus. Where they differed was in terms of fidelity/objectivity and cost efficiency. EWT is a partial measure since it excludes on-board delay whereas CJTD is a total combined measure of ‘at stop’ and ‘on bus’ delay. EWT is also only appropriate for high frequency services, whereas CJTD is applicable to low and high frequency services.

In terms of data collection, EWT is more efficient at the moment since it only requires information at unlinked bus stops whereas CJTD requires stop to stop bus times which are more far more onerous to collect. Only when travel times are provided automatically will CJTD become feasible. A measure has been developed for rail in Sydney to take advantage of the introduction of vehicle devices but for buses only one half of the fleet has had automatic vehicle devices installed.

Table 2: Summary Assessment of Reliability Measures

#	Description	Ease of Understanding	Customer Focus	Fidelity & Objectivity	Cost/Effort Efficiency	Overall Rating	Score Rating	Rank
1	% Buses Cancelled	High	Low	Low	High	Medium	4	4
2	% Departing On-time	High	Low	Medium	Medium	Medium	4	4
3	% Arriving On-time	High	Low	Medium	Medium	Medium	4	4
4	Excess Waiting Time	High	High	Medium	Medium	High	6	1
5	Average Lateness	High	Low	Medium	Medium	Medium	4	4
6	Variability Measures	Low	Low	Low	Medium	Low	1	9
7	Reliability Buffer	Low	Medium	Low	Low	Low	1	9
8	Passenger Ratings	Medium	High	Medium	Low	Medium	4	4
9	Customer Complaints	High	High	Low	Medium	Medium	5	3
10	Customer Delay	High	High	High	Low	High	6	1

Traditionally, there has been an emphasis amongst transport operators to collect statistics on operational performance such as % Buses Cancelled, % Departing or Arriving On-time. They remain popular because the statistics are easy to understand and the data is relatively easy to collect especially if ‘self-reported’ by drivers. Moreover, historical trends have also usually been established.

Independent bus stop reliability surveys using fieldworkers similar to the ITSRR surveys of 2004 and 2005 that have large enough bus stop sample sizes and cover long enough periods of time are expensive to design, implement and report. The resultant statistics should pass the objectivity test and facilitate comparisons across routes and between operators. However, they measure outcomes for buses rather than outcomes for customers although it should be possible to weight observations with passenger load data.

The % Buses Cancelled and % Departing or Arriving On-time measures also have low fidelity. Each is a binary measure (for example, a bus was either cancelled or not cancelled). There is no account taken of the extent to which a service was not on-time. They are also partial measures for customers since they focus on 'at stop' delay and not 'on bus' delay. In summary, these measures are products of an earlier time when automated data capture systems did not exist.

As regards the statistical variability measures of reliability such as headway adherence, they may be useful statistical measures for operators and transport planners but are likely to be less well understood by the public.

The reliability buffer derives from highway engineering work where the emphasis is on road travel times. It is therefore than the other objective measures in focussing on 'on-bus' than 'at-stop' delays. By combining the average and the 95th percentile, the aim is to help passengers plan their departure time. Again, however the measure is unlikely to be well understood.

Passenger ratings are 'perception' based and necessarily subjective. Objective measures would still usually be required to explain the ratings obtained. In this respect, rating surveys should be regarded as one aspect of service monitoring, albeit able to cover a range of service attributes not just reliability, and providing a comparator for objective measures of service reliability.

Passenger complaints are also a useful indicator of service reliability and other aspects of service provision. Complaints result, in most cases, from personal experiences *"My bus has been late three days in a row"*. There is therefore the issue in measuring severity and there is also the need for complaints to be validated from time to time. Where bus services are provided by more than one company, a common system for handling customer complaints is required since the number of complaints will be a related to how easy it is to lodge them.

Measures such as EWT and Average Lateness can, in effect, be calculated for every customer and as such they should be strongly customer-focussed measures. They should also be easily understandable although there is the concern that when presenting in minutes, the concept of variability is lost. The main limitation of EWT is that it is strictly only relevant for high frequency bus routes advertised on a headway basis (e.g. buses run every five minutes between 8.30 and 9.00 am) for which passengers will turn up at the bus stop at random. The measure is not strictly relevant for infrequent services. The other limitation of EWT is that it does not include travel time delay to passengers on the bus.

Customer Journey Time Delay can be calculated for frequent and infrequent services and includes both 'at stop' and 'on bus' delay. The measure has been successfully calculated for Sydney rail passengers. For bus, as mentioned above, the issue is the availability of suitable data.

For rail, the introduction of improved transport control systems has been driven by a need for greater safety and / or greater operational reliability. A by-product of these control systems has been the capture of detailed information on both planned services and actual services. However to date, the availability of detailed information has not significantly influenced the range of measures used by transport operators to assess service delivery. The aforementioned TLS captures detailed information on the movements of CityRail services and, while able to measure operational performance such as % Arriving On-time, it also allows, the calculation of EWT and CJTD. In effect, it was the availability of automatic data that has permitted the calculation of the new measure of customer delay.

With respect to buses the Roads and Maritime Services (RMS) has been developing PTIPS (Public Transport Information and Priority System). Each bus has an on-board GPS unit which transmits to a central location the position of the bus. Also transmitted via data communications are service identifiers such as the route number and which particular trip the bus is operating. This allows PTIPS to effectively compare the actual performance of a bus with the scheduled performance. It is the bus equivalent of CityRail's TLS and as such can be developed to calculate EWT and CJTD. At this time well over half the buses operating route services in the Sydney Metropolitan Area are PTIPS-enabled.

4. Conclusions

Traditionally, there has been an emphasis amongst transport operators to collect statistics on operational performance such as % Buses Cancelled, % Departing On-time and % Arriving On-time. These 'operational' measures remain popular because the statistics are easy to understand and the data is relatively easy to collect especially if 'self-reported' by drivers. In addition, historical trends have usually been established and, as many operators use them, there is within industry comparability.

This paper has reviewed these operationally based statistics with seven other indicators of urban bus service reliability using an assessment framework and an assembly of evidence drawn from a study undertaken for the NSW Independent Transport Safety and Reliability Regulator (ITSRR) in 2006. The original 2006 review concluded that excess waiting time (EWT) was the preferred measure for high frequency bus services and the percentage of buses running on-time the preferred indicator for low frequency services with little to choose between measuring arrival versus departure times.

This paper has updated the original 2006 review to take account of the developments in the automatic collection of travel time data on buses and trains. For rail in Sydney, the incorporation of control systems that record the time of day at which trains stop at stations has enabled a new measure of customer journey time delay (CJTD) to be developed that incorporates 'at station' and 'on train' delays that can be calculated for both frequent and infrequent services.

When considered alongside the other nine reliability indicators in the evaluation framework, the new CJTD measure ranked equal top with EWT. Both measures were considered easy to understand and have a high customer focus. Where they differed was in terms of fidelity/objectivity and cost efficiency. EWT is a partial measure since it excludes on-board delay whereas CJTD is a total combined measure of 'at stop' and 'on bus' delay. EWT is also only appropriate for high frequency services whereas CJTD is applicable to low and high frequency services.

In terms of data collection, EWT is more feasible since it only requires information at unlinked bus stops whereas CJTD requires stop to stop bus times which are far more onerous to collect. However with automatic collection of travel times, CJTD should become a practically feasible measure for bus as has already been demonstrated for rail in Sydney.

Furthermore, future advances in automatic travel time and trip data collection should offer the capability to focus and fine-tune reliability measures even more towards the individual customer experience.

5. References

Achterstraat, P. (2010). “Improving the Performance of Metropolitan Bus Services NSW Transport and Infrastructure” NSW Auditor-General’s Report Performance Audit. Sydney Australia.

Association of Train Operating Companies (2002). Passenger Demand Forecasting Handbook, UK.

Balcombe, R., R. Mackett, N. Paulley, J. Preston, J. Shires, H. Titheridge, M. Wardman and P. White (2004). The Demand for Public Transport; A Practical Guide, Transport Research Laboratory UK.

BTS (2011). "2011 Transport Customer Survey - Customer Satisfaction with Public Transport Services",. Sydney, NSW, report by Bureau of Transport Statistics, NSW 2011, .

Bullock, P., Q. Jiang and P. R. Stopher (2005). " “Using GPS Technology to Measure On-time Running of Scheduled Bus Services” " Journal of Public Transportation **Volume 8, Number 1, 2005.** .

Department of Transport (2008). Track Record - Quarterly Performance Bulletin. Melbourne, Australia, Victorian Government.

Douglas Economics (2006). Monitoring Bus Reliability – A Review of ITSRR Surveys. Sydney, NSW, Report to Independent Transport Safety & Reliability Regulator.

Douglas, N. and G. Karpouzis (2006). Valuing Rail Service Attributes through Rating Surveys. Australasian Transport Research Forum.

FHA (2004). “Traffic Congestion & Reliability – Trends and Advanced Strategies for Congestion Mitigation” US Department of Transportation Federal Highway Administration.

Frumin, M. (2008). Oyster-Based Performance Metrics for the London Overground. London, UK, Transport for London. **Available at http://frumin.net/img/oyster-metrics/20081028%20Oyster-Based%20Performance%20Metrics.pdf (last accessed 26-05-2012)**.

Howard, B. (1995). “Bus Route Evaluation Standards”. TCRP Synthesis of Transit Practice 10, . National Academy Press. Washington DC USA.

ITSRR (2004). Review of on-time running of CityRail services. Sydney NSW, Independent Transport Safety and Reliability Regulator.

ITSRR (2008). Best Practice in Procurement of Commuter Rail Services. Sydney NSW, Independent Transport Safety and Reliability Regulator.

ITSRR (2009). "Survey of Sydney Metropolitan Bus Users 2009",. Sydney, NSW, report by Independent Transport Safety and Reliability Regulator, September 2009.

ITSRR (2010). “Transport Reliability Report 08-09” Sydney NSW, Independent Transport Safety and Reliability Regulator.

Kittelson & Associates, KFH Group, Parsons Brinkerhoff Quade and Douglas Inc and K. JHunter-Zaworski (2003). Transit Capacity and Quality of Service Manual - 2nd Edition. Transit Cooperative Research Program TCRP Washington DC.

Kittleson & Associates, U., LKC Consulting Services Inc, MORPACE International Inc, Queensland University of Technology, Yuko Nakanishi, (2003). ‘A Guidebook for Developing a Transit Performance-Measurement System’ Transit Cooperative Research Program TCRP Report 88, Transportation Research Board Washington DC 2003. .

Mazloumi E, Currie G and Sarvi M (2008). Assessing Measures of Transit Travel Time Variability Using AVL Data. Transportation Research Board Meeting Washington DC USA.

Mezghani, M. (2008). “Study on Electronic Ticketing in Public Transport”, , report for the European Metropolitan Transport Authorites

Ministry of Transport NSW (2006). “Submission to the Independent Pricing & Regulatory Tribunal on Bus Fares for 2007”. . Sydney, NSW, Independent Pricing & Regulatory Tribunal, NSW.

MVA (2007). Passengers' Priorities for Improvements in Rail Services, . London, UK, Report for Passenger Focus

Nakanishi, Y. J. (1997). " “Bus Performance Indicators: On-time Performance and Service Regularity”. In " Transportation Research Record 1571: pp3-13.

PCAC (2011). “Minutes Matter – A review of performance metrics at the MTA” New York USA, research report by the Permanent Citizens Advisory Committee to the MTA.

Quiroga (1997) “An Integrated GPS-GIS Methodology for Performing Travel Time Studies’ PhD Dissertation Louisiana State University Baton Rouge, LA.

Strathman, J. G., K. J. Dueker, R. L. Kimpel, R. L. Gerhart, K. Turner, P. Taylor, S. Callas and D. Griffin (2000). "“Service Reliability Impacts of Computer-aided Dispatching and Automatic Vehicle Location Technologies: a Tri–Met Case Study”." Transportation Quarterly **54 (3)**.

Sydney Buses (2006). “State Transit Service Performance” Sydney, NSW, State Transit Authority of NSW.

Sydney Ferries (2012). “How We Are Charting” http://www/sydneyferries.info/performance/how-we-are-charting.htm Last Accessed 04-2012.

Transport for London (2012). “London Buses Quality of Service Indicators – Route Results for London Buses Services” London, UK. **Fourth Quarter 2011/12 7th Jan 2012-31st March 2012”**

Trompet, M., X. Lui and D. J. Graham (2011). "“Development of Key Performance Indicator to Compare Regularity of Service Between Urban Bus Operators”." Transportation Research Record: Journal of the Transportation Research Board, **No. 2216**: pp. 33–41.