Research in Transportation Economics xxx (2014) 1–9

Contents lists available at ScienceDirect

Research in Transportation Economics

journal homepage: www.elsevier.com/locate/retrec

Incorporating service reliability in public transport design and performance requirements: International survey results and recommendations

Niels van Oort [a], [b], *

[a] Department of Transport and Planning, Delft University of Technology, The Netherlands
[b] Goudappel Coffeng Mobility Consultants, The Hague, The Netherlands

ARTICLE INFO

Article history:
Available online xxx

JEL classification:
R4

Keywords:
Public transport
Service reliability
Monitoring
Performance requirements

ABSTRACT

Although public transport passengers consider service reliability a key quality aspect, actual services are often not perceived as reliable. To gain insights into how authorities deal with (improving) service reliability, an international survey was performed, showing that very little attention is paid to service reliability during the design of the network and of the timetable. It also illustrates that there is little consistency in approaches. In addition, a second survey was performed, showing how Dutch authorities deal with service reliability in relation to concession requirements and incentive regimes. The main findings are that consistency is lacking on this topic and that minimum attention is paid to passenger impacts of unreliability in concession requirements. This results in services that do not match the (implicitly) required level of service reliability. These surveys also revealed that there is no consistency in the definition of service reliability. For instance, traditional indicators focus on vehicles instead of passengers. By using an alternative reliability indicator, additional travel time, we demonstrated that traditional indicators lead to wrong indications. Based on our findings, we present recommendations to improve concession requirements as well as to improve the design of networks and of timetables, both aiming at enhanced service reliability.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

Service reliability is a key quality indicator for public transport and has become increasingly important over the last years (European Commission, 2011; Levinson, 2005; OECD/International transport forum, 2010). However, in most countries, passengers do not perceive actual services as reliable. In Van Oort (2011), a framework as well as cost-effective design instruments are presented to improve the level of service reliability from a passenger perspective. The key to developing potential improvement instruments was analysing Big Data of both vehicle performance (AVL data, see for instance Hickman, 2004) and passenger flows (APC data, see for instance Pelletier, Trepanier, & Morency, 2011).

This papers aims to gain insights into how public transport authorities deal with (improving) service reliability and the design instruments of Van Oort (2011). To achieve this objective, an international survey was performed, followed by a second survey in The Netherlands, showing how public transport authorities deal with service reliability in concession requirements and in incentive regimes. In this paper, the results of these two surveys are presented. Furthermore, recommendations are suggested to improve concession requirements as well as to improve the design of networks and of timetables, both aiming at enhanced service reliability.

The remainder of this paper is structured as follows: Section 2 will describe service reliability from a passenger perspective and illustrates how design choices may enhance the actual level. Section 3 will present the two surveys and their results concerning design. Since measuring service reliability proved to be an important topic, Section 4 will deal with that topic, from a theoretical as well as practical (i.e. survey results) perspective. Finally, results are discussed and acknowledgements and references are provided.

2. Service reliability and design

Service reliability is the certainty of service aspects compared to the schedule as perceived by the user and is one of the most

---

* Department of Transport and Planning, Delft University of Technology, P.O. Box 5048, 2600 GA, The Netherlands. Tel.: +31 6 15908644.
E-mail address: N.vanOort@TUDelft.nl.

http://dx.doi.org/10.1016/j.retrec.2014.09.036
0739-8859/© 2014 Elsevier Ltd. All rights reserved.

Please cite this article in press as: van Oort, N., Incorporating service reliability in public transport design and performance requirements: International survey results and recommendations, *Research in Transportation Economics* (2014), http://dx.doi.org/10.1016/j.retrec.2014.09.036

important quality aspects of public transport (Van Oort, 2011). Actual vehicle trip time variability (i.e. service variability) affects service reliability and passenger travel time. The impacts of unreliable services on passengers are:

- average travel time extension;
- increased travel time variability;
- a lower probability of finding a seat in the vehicle.

Several researches present quantification methods of the costs of these elements, thereby proving that these might be substantial (see for instance Li, Hensher, & Rose, 2010; Rietveld, Bruinsma, & van Vuuren, 2001; Schmöcker, Fonzone, Shimamoto, Kurauchi, & Bell, 2011).

Literature shows that, for urban public transport, serious attention is given to ways to improve service reliability at the operational level (see for instance Muller & Furth 2000; Osuna & Newell, 1972). One of the issues during operations for instance is increased dwell time due to poor service regularity. Potential service reliability improvement instruments also exist during the public transport design stages. Much research is already available on the planning instruments of priority at traffic lights, exclusive lanes and synchronisation. The implementation of bus lane schemes and traffic signal priority are the most used solutions in this field. Both Ceder (2007) and the Transit Capacity and Quality of Service Manual (TRB, 2003) present the different methods and effects. More detailed research on the impacts of priority and bus lanes may be found in Chang, Collura, Dion, and Rakha (2003), Shalaby (1999) and Kimpel, Strathman, Bertini, and Callas (2005). Similar to priority, Ceder (2007) gives an overview of the issues which need to be considered in synchronisation. More detailed and applied research is for instance done by Wong and Wilson (2006).

In literature, five additional potential design instruments are found. These instruments proved to be capable of increasing the level of service reliability. Simulation studies showed that these instruments may be cost-effectively implemented and that they may attract 5–15% more passengers (Van Oort, 2011). Below, these instruments are presented. In the next section, the survey results are shown, indicating how these instruments are applied in practice. At the strategic level, these instruments are (the references provide more detailed information on these instruments and their impacts):

- Terminal design (Van Oort & Van Nes, 2010);
  The configuration and number of tracks and switches at the terminal determines the expected vehicle delay and thus service reliability.
- Line length (Van Oort & Van Nes, 2009a);
  The length of a line is often related to the level of service variability and thus service reliability.
- Line coordination (Van Oort & Van Nes, 2009b).
  Multiple lines on a shared track may offer a higher level of service reliability than one line (assuming equal frequencies).

The following instruments may be applied at the tactical level (the references provide more detailed information on these instruments and their impacts):

- Trip time determination (Van Oort, Boterman, & Van Nes, 2012);
  In long-headway services, scheduled vehicle departure times at the stop, derived from scheduled trip times, determine the arrival pattern of passengers at their departure stop. Adjusting the scheduled trip time may affect the level of service reliability and passenger waiting time.
- Vehicle holding (Van Oort, Wilson, & Van Nes, 2010).
  Holding early vehicles reduces driving ahead of schedule and increases the level of service reliability. The design of the schedule affects the effectiveness of this instrument.

The terminal design instrument relates to (new) rail lines with tail tracks as terminal or short-turning facilities. For high-frequency, distributed lines, compact tail tracks with double crossovers directly after the stop are recommended. Concerning (new) lines with a clear break point in passenger pattern, it is recommended to split the line or to apply holding points. For long-headway services, the 35-percentile value for scheduled trip time is proposed. If parts of lines are very crowded, investigating the effects of coordination is suggested.

Since these instruments proved to be potential, we investigated in a survey how these instruments are applied in practice, presented by the next section.

3. Two surveys on public transport service reliability

To gain insight into how public transport authorities and operators currently deal with service reliability, we performed two surveys. The first one focused on service reliability in relation to design of public transport (specifically focussing on the five instruments of the previous section). The second one focused on the role of service reliability in tender requirements. Both surveys are described in the following sections in more detail. Sections 3.3 and 4 present the main findings of both surveys.

3.1. *International survey*

The objective of the international survey was to learn about reliability and design topics in several cities in different countries, focussing on urban and regional public transport. The survey also demonstrates how public transport operators quantify service reliability in practice, and provides insights into design guidelines that might affect service reliability. The survey consisted of a questionnaire that was sent all over the world. Responses were received by almost 30 authorities and operators, mainly Western countries. Table 1 shows the respondents.

3.2. *Dutch survey on tender requirements*

Since measuring and quantifying service reliability proved to be an important issue in the international survey, we performed a more detailed second one. This one was held in 2012 in the Netherlands under supervision of KPVV, the Dutch Knowledge Centre on Transport. The survey consisted of two parts: desk research and interviews. The desk research was performed by analysing a random selection of recent tender documents, with

Table 1
Participating cities and systems in international reliability survey.

City	PT type	City	PT type	City	PT type
Amsterdam	Metro, tram, bus	Gothenburg	Tram	Rouen	Tram, bus
Barcelona	Metro, bus	Halle	Tram, bus	Salt Lake City	Light Rail
Berlin	S-Bahn, tram	Hong Kong	Light rail	Stockholm	Metro, bus
Brussels	Tram	Lolland	Bus	Stuttgart	Rail
Chicago	Metro, bus	London	Tram, bus	Santa Cruz de Tenerife	Tram
The Hague	Light rail, tram, bus	Milano	Bus, tram	Vienna	Metro, tram, bus
Dresden	S-Bahn, tram	Minneapolis	Bus	Zurich	S-Bahn
Dublin	Tram	Rotterdam	Metro, tram, bus		

Please cite this article in press as: van Oort, N., Incorporating service reliability in public transport design and performance requirements: International survey results and recommendations, *Research in Transportation Economics* (2014), http://dx.doi.org/10.1016/j.retrec.2014.09.036

specific attention to service reliability. Fig. 1 shows the regions of which the tender documents were investigated. In total, we analysed 22 tender documents. Both rail and road bound transport were part of the selection.

In addition to the desk research we performed interviews with a selection of twelve Dutch public transport authorities. These are:

- Province of Flevoland;
- Province of Friesland;
- Province of Gelderland;
- Province of Limburg;
- Province of Noord-Holland;
- Province of Overijssel;
- Province of Utrecht;
- Province of Zeeland;
- Region of Arnhem and Nijmegen;
- Region of Eindhoven;
- Region of Groningen and Drenthe;
- Region of Twente.

The interview topic was how public transport authorities deal with service reliability (improvements) in general, with specific regard to performance requirements and monitoring regimes.

Both surveys showed that there is little attention paid to service reliability during the design of the network and of the timetable. In addition, it illustrated that little consistency exists in approaches. In this paper we illustrate the consequences of this inconsistency. In the next section we show the results of the international survey related to design and the five design instrument introduced in Section 2. The results of quantifying and measuring service reliability will be presented in Section 4.

3.3. Survey results related to design

Our international survey showed that not all of the instruments presented in Section 2 are common use yet. None of the participants considered the impact of the length of the line on service reliability, although longer lines may have a negative impact on service quality (as illustrated by Van Oort & Van Nes, 2009a). With regard to trip time determination, most authorities and operators used large percentile values to determine trip time (50% or higher). Both surveys illustrated that there is little attention to the impact of tactical design choices on operational quality. An important issue is lack of requirements concerning trip time determination. Although there is a strong relationship between the method of trip time determination and service reliability, most investigated Dutch tender documents only mentioned that trip times should be realistic. However, Van Oort et al. (2012) proved that too much buffer in trip times has a large impact on passenger travel times (due to early departures) and that too little buffer time may create many delays. Holding is a popular instrument (78% of the participants apply holding), but little attention is paid to the relation to trip time determination as well (while Van Oort et al., 2010 demonstrated a direct relationship). About 70% of the participants apply coordination and in case of tail track terminals, about 35% of the designs have double crossovers before the platform, which is the optimal configuration concerning service reliability (Van Oort & Van Nes, 2010).

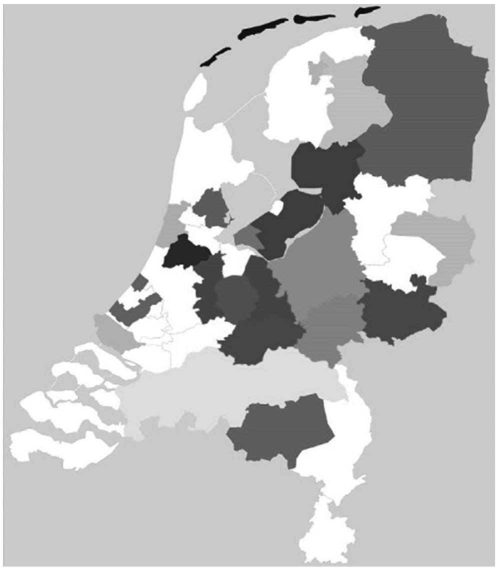

Fig. 1. Investigated regions in the Netherlands (separately indicated in grey).

New Dutch data sources as smartcards (APC) and "GOVI" (Dutch AVL data, see for instance Van Oort, Sparing, & Brands, 2013) are considered very promising by the participants as source to improve reliability. Big Data in public transport offers great opportunities to analyse past performance and find potential improvements.

The survey showed that not all of the five mentioned instruments are applied yet. One conclusion is that to make optimal design choices, proper indicators are required. Concerning service reliability, the traditional indicators are not considered helpful in that way. The next section will elaborate on that topic.

4. Survey results related to service reliability measurement

In order to improve service reliability, it is essential to monitor and predict the level of service reliability of a public transport system. For this we need proper indicators. The commonly used indicators which are supposed to express reliability do not completely focus on the impact on passengers of service reliability. In fact, they focus more on service variability of the system than on the actual impacts on passengers. This section presents the traditionally used indicators and describes an alternative indicator that enables enhanced quantification of service reliability.

4.1. Traditionally used indicators

Given the stochastic nature of public transport operations, statistical measures such as standard deviation or percentiles are logical indicators for service reliability. A typical example is the coefficient of variation of headway, as shown by Equation (1). This indicator may relate to an aggregate characteristic of a public transport line, or a branch served by a set of public transport lines. Equation (1) shows the coefficient of variation of actual headways per stop, but in practice expressing this indicator on line level by calculating the average value over the stops, is also common. This way, the number of passengers per stop is neglected.

$$CoV\left(\tilde{H}_{l,j}^{act}\right)=\frac{StD\left(\tilde{H}_{l,j}^{act}\right)}{E\left(\tilde{H}_{l,j}^{act}\right)} \tag{1}$$

where:

$CoV(\tilde{H}_{l,j}^{act})$ = coefficient of variation of actual headways of line l at stop j
$\tilde{H}_{l,j}^{act}$ = actual headway of line l at stop j
$StD(\tilde{H}_{l,j}^{act})$ = standard deviation of actual headways of line l at stop j

Please cite this article in press as: van Oort, N., Incorporating service reliability in public transport design and performance requirements: International survey results and recommendations, *Research in Transportation Economics* (2014), http://dx.doi.org/10.1016/j.retrec.2014.09.036

$E(\tilde{H}^{act}_{l,j})$ = expected headway of line l at stop j

In practice, however, the use of purely statistical measures is limited. Commonly used indicators focus either on punctuality, the extent to which the scheduled departure times are met, or on regularity, the variation in the headways.

From the perspective of the production process, the percentage of trips performed within a predefined bandwidth, is a useful reliability indicator. Equation (2) expresses this type of indicator for average departure deviation for a complete line. Observed data is used to determine the relative frequency of deviations within a bandwidth. This indicator represents to which extent the production process requirements are met. The next section will present actual used values of $\delta^{\min}$ and $\delta^{\max}$. Obviously, these values are of great influence on the level of service reliability calculated.

$$P_l = \frac{\sum_{j=1}^{n_{l,j}} \sum_{i=1}^{n_{l,i}} P_{l,i,j}\left(\delta^{\min} < \tilde{D}^{act}_{l,i,j} - D^{sched}_{l,i,j} < \delta^{\max}\right)}{n_{l,i}{*}n_{l,j}} \tag{2}$$

where:

P_l = relative frequency of vehicles on line l having a schedule deviation between $\delta^{\min}$ and $\delta^{\max}$
$P_{l,i,j}$ = relative frequency of vehicle i on line l having a schedule deviation between $\delta^{\min}$ and $\delta^{\max}$ at stop j
$\tilde{D}^{act}_{l,i,j}$ = actual departure time of vehicle i on stop j on line l
$D^{sched}_{l,i,j}$ = scheduled departure time of vehicle i on stop j on line l
$\delta^{\min}$ = lower bound bandwidth schedule deviation
$\delta^{\max}$ = upper bound bandwidth schedule deviation
$n_{l,i}$ = number of trips of line l
$n_{l,j}$ = number of stops of line l

Punctuality may also be defined as the (average) deviation from the timetable at a specific stop, a set of stops, or for all stops of a line. The latter is shown by Equation (3) (Hansen, 1999).

$$p_l = \frac{\sum_{j}^{n_{l,j}} \sum_{i}^{n_{l,i}} \left| \tilde{D}^{act}_{l,i,j} - D^{sched}_{l,i,j} \right|}{n_{l,j}{*}n_{l,i}} \tag{3}$$

where:

p_l = average punctuality on line l

Please note that this formulation has an important shortcoming. It does not indicate whether vehicles depart too early or too late, which has a large impact on passenger waiting time. If only a set of stops is considered, the location of the stops may be of influence.

Irregularity is used to express headway deviations. Hakkesteegt and Muller (1981) introduced the PRDM (Percentage regularity deviation mean), which shows the average deviation from the scheduled headway as a percentage of the scheduled headway. The calculation of the PRDM is shown in Equation (4). This equation shows the calculation of the PRDM per stop. Taking into account all the stops, a calculation of the PRDM for the total line is also possible.

$$PRDM_{l,j} = \frac{\sum_{i} \left| \frac{H^{sched}_{l,i} - \tilde{H}^{act}_{l,i,j}}{H^{sched}_{l,i}} \right|}{n_{l,j}} \tag{4}$$

where:

$PRDM_{l,j}$ = relative regularity for line l at stop j
$H^{sched}_{l,i}$ = scheduled headway for vehicle i on line l
$\tilde{H}^{act}_{l,i,j}$ = actual headway for vehicle i on line l at stop j
$n_{l,j}$ = number of vehicles of line l departing at stop j

All of the presented measures focus purely on characteristics for the supply side, although it should be noted that indicators for punctuality and regularity are linked with assumptions on the arrival pattern of travellers, i.e. arrivals based on the timetable and uniformly distributed arrivals respectively. More important is the fact that these measures make no distinction between stops having a high demand or a low demand. Punctuality and regularity have a strong influence on waiting time and are thus most important for stops having large numbers of passengers boarding the vehicles. Furthermore, these indicators do not quantify the impact the variability has on travellers, such as the extra travel time as discussed in the previous section. The next section will present the results of the surveys concerning the use of these indicators.

4.2. *Indicators in practice*

The previous section dealt with quantifying service reliability. More than one method/indicator is used in practice to present the level of service reliability. Both in theory and practice, analysts tend to focus on supply-side indicators which do not illustrate the actual service reliability, but rather show the output variability of the system. Most applied measures focus on departure time deviations. Mostly, early and late vehicles are treated as the same. The departure time deviations may be calculated per line or network, including all or only a few stops.

Another way to express schedule adherence is the percentage of vehicles' schedule deviations within a certain bandwidth. Equation (2) showed how to calculate this indicator (in which $\delta^{\min}$ and $\delta^{\max}$ represent the lower and upper bound respectively). This method is very common in heavy railways. The Dutch Railways, for instance, used to periodically present the number of trains departed not later than 3 min from 32 main stations in the Netherlands until 2010 (i.e. $\delta^{\max} = 3$ min). Most heavy railway companies in Europe use 5 min as a maximum (Landex & Kaas, 2009; Schittenhelm & Landex, 2009), as the Dutch railways do currently. In the U.S., even 30 min delay is considered being on time (Bush, 2007). Among the urban public transport industry, sometimes the bandwidth has a lower boundary value as well (i.e. $\delta^{\min}$), which means that driving ahead of schedule is considered explicitly. For example, vehicles are considered punctual when they depart between 0 and +5 min compared to the schedule (Nakanishi, 1997). Of the participating cities in the international survey, 74% use a bandwidth to quantify and analyse schedule adherence, while 21% use the average punctuality. The results of the survey showed that only London has another way of measuring the difference between schedule and operations, being excess journey time.

Besides different indicators (average punctuality, bandwidth punctuality and regularity) the boundaries of the bandwidth are not uniform (see Fig. 2). This figure shows the different values used by the participants of the international survey, where every line corresponds to the boundaries of one city or system. It is shown that three cities do not even use a lower boundary value (indicated as −5 min). The maximum boundary value ranges from +1 to +6 min. These differences in bandwidth obviously have a large impact on the percentage of on-time vehicles. Setting the requirement for excess variability thus determines the quality of operations. If a broad bandwidth is set, excess variability will be small for instance. Fig. 3 shows the results for the Dutch survey. One major difference with the international results is that there is a lower boundary (no early departing) in almost all cases. On the other hand, the maximum boundary is higher in some cases.

Please cite this article in press as: van Oort, N., Incorporating service reliability in public transport design and performance requirements: International survey results and recommendations, *Research in Transportation Economics* (2014), http://dx.doi.org/10.1016/j.retrec.2014.09.036

ARTICLE IN PRESS

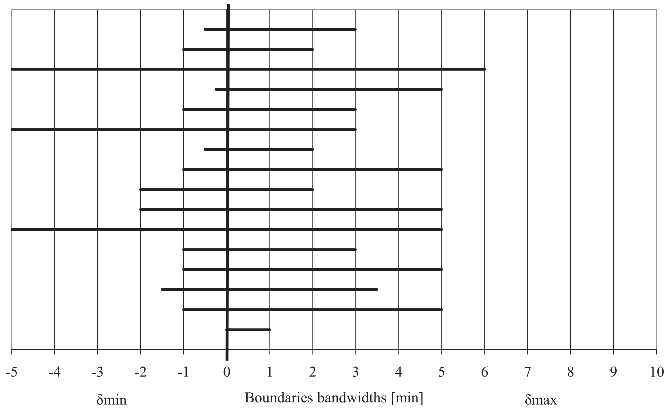

Fig. 2. International survey: boundaries of bandwidths applied in sample cities, to measure departure reliability at stops.

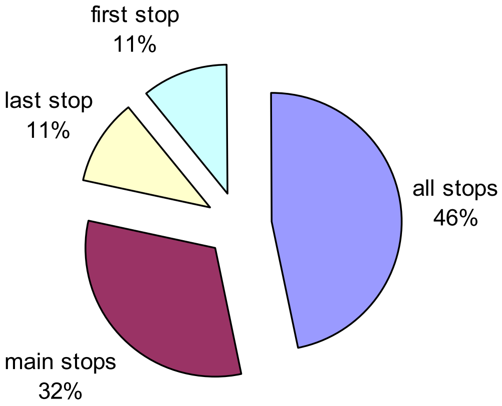

Fig. 4. Locations used to measure service reliability (i.e. departure time deviations) (international survey).

However, from a passenger perspective, late vehicles tend to affect travel times less than early vehicles. With regard to penalising unpunctual vehicles, in 18% of the investigated tender documents, a penalty regime was applied.

With regard to penalising a service quality level for being too low, not all experiences were satisfactory. First of all it is hard to precisely define an ambitious yet achievable level of service. Authorities tend to set too high standards, from the perspective of operators. When a penalty is proposed or applied, much discussion starts on how the data is achieved and processed. In addition, it is very important to distinguish who is responsible for which part of unreliability: as shown in Van Oort (2011), several sources together create variability and unreliability. Some of them are under the responsibility of the operator and some under the public transport authority and/or infrastructure manager.

Besides differences in indicators and boundaries, locations of measuring service reliability differ among the participating cities as well. Sometimes, only departure at the terminal is considered or just the main stops. Fig. 4 shows the response on the question of where to measure service reliability (i.e. departure time deviations) of the international survey. In the Dutch survey, we find that in 18% of the investigated concessions punctuality was measured at the first stop, in 27% at the last stop and in 23% at the main transfer points.

These results show that there is no uniform method applied in practice to measure service reliability, although this quality aspect is considered very important. Besides, the focus in practice is mainly on the supply side of public transport. Only Transport for London uses indicators showing the effects of unreliability for passengers, being excess journey time. These survey results support the statement of Van Oort and Van Nes (2009b) to introduce a new indicator for reliability, namely the average additional travel time per passenger. In literature, the need for a more passenger-focused indicator is recognised as well (e.g. Frumin, Uniman, Wilson, Mishalani, & Attanucci, 2009; Landex & Nielsen, 2006; Mazloumi, Curry, & Majid, 2008).

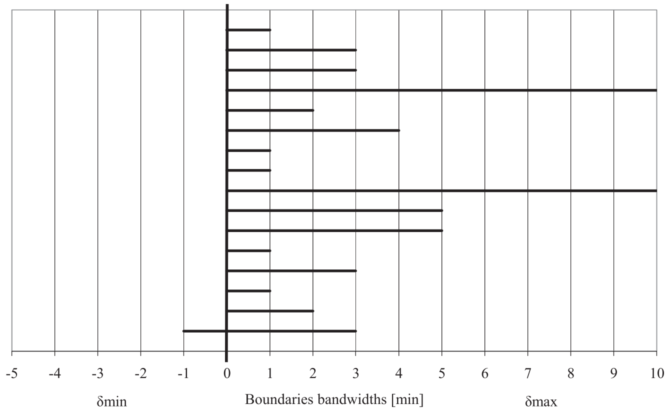

Fig. 3. Dutch survey: boundaries of bandwidths applied in sample cities, to measure departure reliability at stops.

4.3. Alternative service reliability indicator

Several traditional quantifications of service reliability, such as punctuality and regularity, have a lack of attention for passenger impacts. Traditional indicators focus too much on the supply side of public transport, which does not allow a proper analysis of passenger effects. To deal with the shortcomings of traditional indicators, Van Oort and Van Nes (2009b) developed a new indicator, being the average additional travel time per passenger. This indicator translates the supply-side indicators, for instance punctuality, into the additional travel time that a passenger on average needs to travel from the origin to the destination stop due to service variability. The average additional travel time may be calculated per stop or per line and it enables explicit consideration of service reliability in cost-benefit calculations, since the level of service reliability may be translated into regular travel time.

In order to calculate the additional waiting time component, two situations have to be distinguished: high frequency transit systems (with random arrivals of passengers at the stop) and low frequency transit systems (with planned arrivals of passengers at the stop).

If passengers arrive randomly, exact departure times and punctuality are not relevant anymore, because passengers do not use a schedule. In that scenario, the additional travel time is calculated using the coefficient of variation (CoV) of the actual headways ($\tilde{H}_{l,j}^{act}$). A generic formulation to estimate the expected waiting time per passenger is given by Equation (5) (Osuna & Newell, 1972; Welding, 1957), according to the following assumptions:

- The examined period is homogeneous concerning scheduled departure times, trip times and headways (for instance rush-hour on working days in a month);
- The passenger pattern on the line is assumed to be fixed;
- All passengers are able to board to the first arriving vehicle.

Please cite this article in press as: van Oort, N., Incorporating service reliability in public transport design and performance requirements: International survey results and recommendations, *Research in Transportation Economics* (2014), http://dx.doi.org/10.1016/j.retrec.2014.09.036

$$E\left(\tilde{T}^{waiting}_{l,j}\right) = \frac{E\left(\tilde{H}^{act}_{l,j}\right)}{2} * \left(1 + CoV^2\left(\tilde{H}^{act}_{l,j}\right)\right) \tag{5}$$

where:

$\tilde{T}^{waiting}_{l,j}$ = passenger waiting time for line l at stop j
$\tilde{H}^{act}_{l,j}$ = actual headway of line l at stop j
$CoV(\tilde{H}^{act}_{l,j})$ = coefficient of variation of actual headways of line l at stop j

If the service is regular, the coefficient of variation equals zero and the average waiting time will be equal to half the headway. In the case of irregular service, the additional waiting time may then be calculated using Equation (6). Assuming no change in the actual vehicle trip times, the total average additional travel time per passenger will be equal to the average additional waiting time per passenger.

$$E\left(\tilde{T}^{Add,waiting}_{l,j}\right) = \frac{E\left(\tilde{H}^{act}_{l,j}\right)}{2} * \left(CoV^2\left(\tilde{H}^{act}_{l,j}\right)\right) \tag{6}$$

where:

$E(\tilde{T}^{Add,waiting}_{l,j})$ = average additional waiting time per passenger due to unreliability of line l at stop j

For low frequency services it is assumed that passengers plan their arrival at the first stop of their trip according to the schedule and therefore another method of calculating additional travel time is necessary. Equations (7) and (8) show this method (Van Oort et al., 2012). Passengers are assumed to arrive randomly within a range of the scheduled departure time minus τ_{early} and plusτ_{late} and if the vehicle departs within this time window it is assumed that passengers do not experience any additional waiting time. Research about empirical values of τ_{early} and τ_{late} is presented in (Van Oort et al., 2012). It is important to note that there is a difference between driving ahead of schedule and driving late. Driving ahead (i.e. departing before the scheduled departure time minusτ_{early}) leads to a waiting time equal to the headway (H^{sched}_l; assuming punctual departure of the successive vehicle). Especially in the case of low frequencies, this leads to a substantial increase in passenger waiting time. Arriving late creates an additional waiting time equal to the delay ($\tilde{d}^{departure}_{l,i,j}$). Just as before, the additional waiting time is first calculated per stop.

$$\begin{cases} \tilde{T}^{Add,waiting}_{l,i,j} = H^{sched}_l & \text{if } \tilde{d}^{departure}_{l,i,j} \leq -\tau_{early} \\ \tilde{T}^{Add,waiting}_{l,i,j} = 0 & \text{if } -\tau_{early} < \tilde{d}^{departure}_{l,i,j} < \tau_{late} \\ \tilde{T}^{Add,waiting}_{l,i,j} = \tilde{d}^{departure}_{l,i,j} & \text{if } \tilde{d}^{departure}_{l,i,j} \geq \tau_{late} \end{cases} \tag{7}$$

$$E\left(\tilde{T}^{Add,waiting}_{l,j}\right) = \frac{\sum_i E\left(\tilde{T}^{Add,waiting}_{l,i,j}\right)}{n_{l,i}} \tag{8}$$

where:

$E(\tilde{T}^{Add,waiting}_{l,i,j})$ = average additional waiting time per passenger due to unreliability of vehicle i of line l at stop j
H^{sched}_l = scheduled headway at line l
$\tilde{d}^{departure}_{l,i,j}$ = departure deviation of vehicle i at stop j on line l
τ_{early} = lower bound of arrival bandwidth of passengers at departure stop
τ_{late} = upper bound of arrival bandwidth of passengers at departure stop
$n_{l,i}$ = number of vehicles i on line l

Based on the average additional travel time per passenger per stop of a line, the average additional travel time per passenger on the complete line is calculated. To do this, the proportion or percentage of boarding passengers per stop is used ($\alpha_{l,j}$), as shown by Equation (9). Please note that using the proportion of passengers makes the indicator independent of the actual number of passengers.

$$E\left(\tilde{T}^{Add,waiting}_{l}\right) = \sum_j \left(\alpha_{l,j} * E\left(\tilde{T}^{Add,waiting}_{l,j}\right)\right) \text{ with } \sum_j \alpha_{l,j} = 1 \tag{9}$$

where:

$\alpha_{l,j}$ = proportion of passengers of line l boarding at stop j

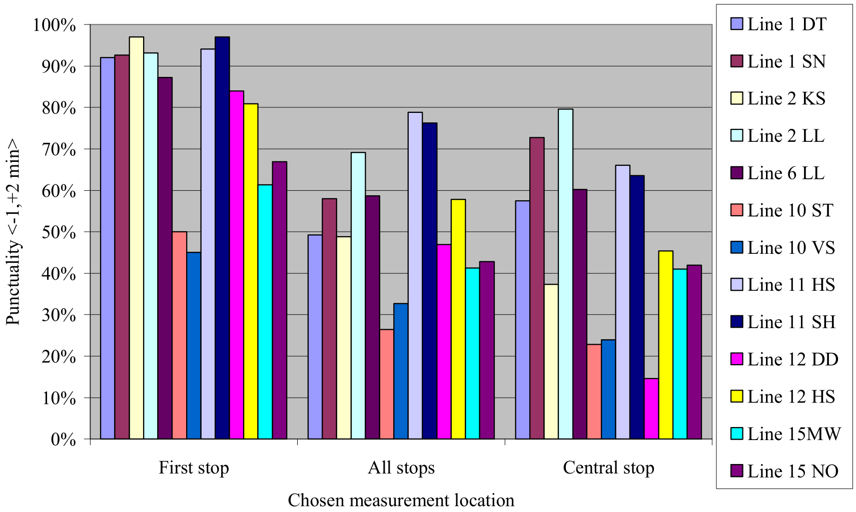

Fig. 5. Punctuality <−1,+2> of tram lines in The Hague using different measurement locations.

Please cite this article in press as: van Oort, N., Incorporating service reliability in public transport design and performance requirements: International survey results and recommendations, *Research in Transportation Economics* (2014), http://dx.doi.org/10.1016/j.retrec.2014.09.036

4.4. *Limitations of service reliability definitions used in practice*

The previous sections showed that there are many methods to illustrate service reliability and that these methods are applied differently in practice. In this section, the impacts of the measurement location and the definition of punctuality are analysed. To show the effects of using these different methods, a case study is conducted using empirical data of tram lines in The Hague, being the 3rd largest city in the Netherlands (about 500.000 inhabitants). All tram lines are analysed and data of rush-hours on

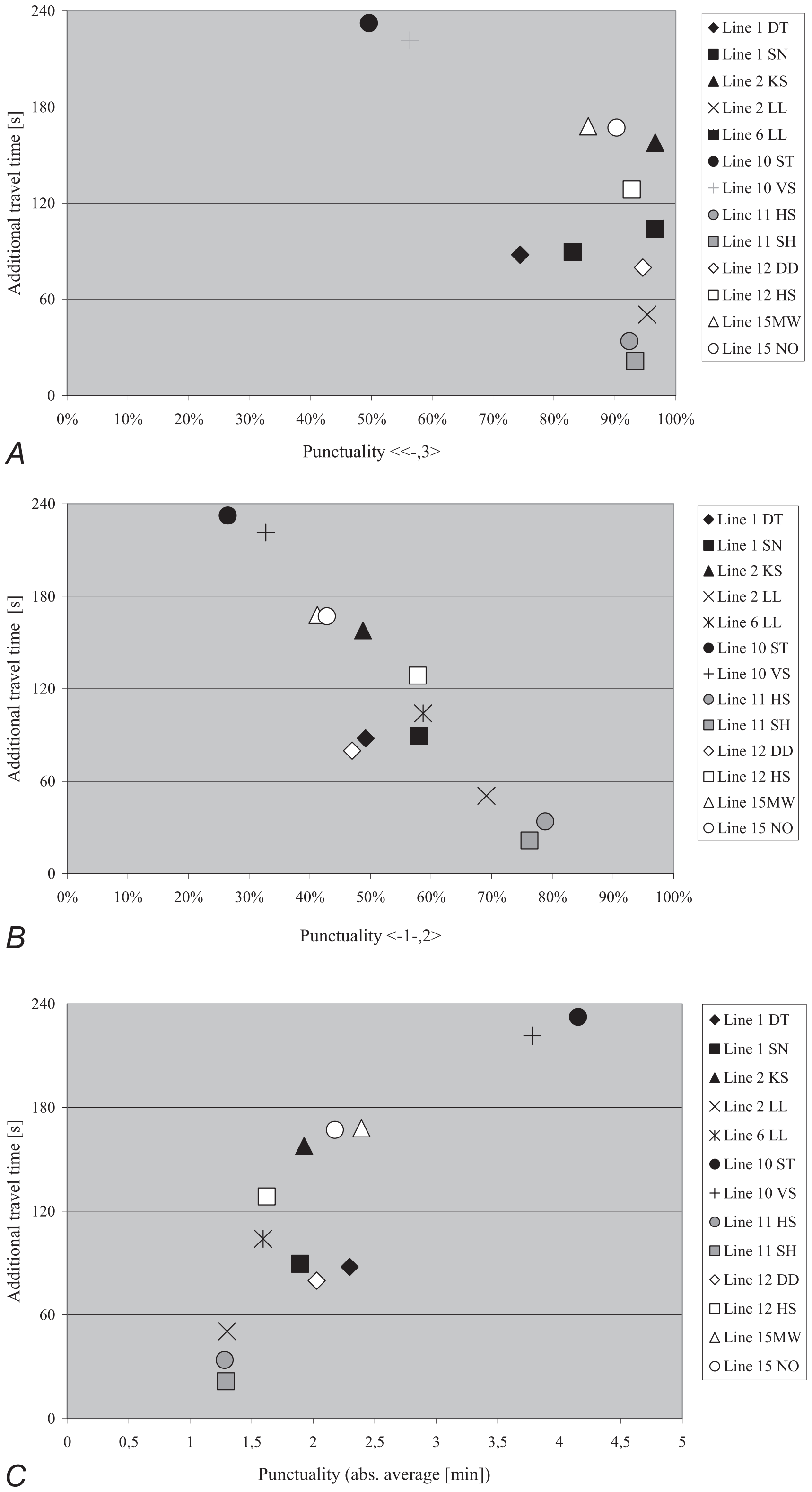

Fig. 6. Calculated additional travel compared to three types of punctuality measurement.

Please cite this article in press as: van Oort, N., Incorporating service reliability in public transport design and performance requirements: International survey results and recommendations, *Research in Transportation Economics* (2014), http://dx.doi.org/10.1016/j.retrec.2014.09.036

working days in April 2007 are used. Fig. 5 shows the impact of different measurement locations on service reliability. This figure illustrates the difference between measuring only at the first stop, at a central stop or at all stops. Fig. 4 already showed that all of these methods are regularly applied in practice. To express service reliability, a bandwidth of timetable deviation of −1 and +2 min is used. The figure shows per tram line the percentage of vehicles departing on the specific stop(s) between these boundary values. It is shown that the different methods do not yield consistent results. The punctual trip percentage per tram depends on the measurement method and the order of tram lines differs per measurement method as well. Line 2 in the direction of KS and line 11 SH prove for example to be the most reliable lines using the first stop measurement, but if only a central stop is investigated, line 2 LL and line 1 SN are more reliable. If all stops are inserted in the calculation, line 11 HS is the most reliable line. This case proves that different methods do not yield comparable results and thus a consistent method is recommended.

Besides the location of measurement, the indicator used is also of great importance and influence. As stated in Section 4.2, punctuality is a supply-focused indicator which is commonly used in urban public transport. The definition of punctuality differs among cities and countries as well, as mentioned in the previous section. In Van Oort and Van Nes (2009b), a new indicator, additional travel time, was introduced. This indicator enables an improved illustration of the level of service reliability; the focus is on the passenger, there is only one definition and it is comparable to travel time.

Fig. 6 A, B and C show a comparison between three definitions of departure punctuality found in the international survey and show additional travel time for actual tram lines in The Hague. The used definitions of punctuality are (calculated for all trips at all stops):

A The percentage of schedule deviations that is less than 3 min late;
B The percentage of schedule deviations which is both less than 2 min late and more than 1 min early;
C The absolute average of the deviation.

Although these figures roughly show a linear relationship between these indicators and the additional travel time, the order of tram lines regarding the highest reliability differs per indicator. For example, line 15 MW has a low reliability using category B (<−1,+2>; only 40%), but a high reliability in category A (≪−,+3>; 85%). Tram lines with many early departures score better on reliability when no lower boundary is used.

Another example of inconsistency is line 2 KS, which has a better reliability than line 1 SN if category B is used. However, the additional travel time per passenger is higher in the first line, so the passenger's experience will not be aligned with the common indicators result. Looking at Fig. 6C, it is illustrated that line 12 DD and 2 KS have about the same value of punctuality, but the average additional travel time per passenger of the latter is about two times higher. The level of service reliability thus depends on the chosen definition. The additional travel time only has one definition and is better suited to addressing the level of service reliability of a specific line or network. This indicator really shows the impact on passengers.

5. Conclusions and recommendations

This paper dealt with service reliability and how public transport authorities deal with this important quality aspect during design, monitoring and tendering of services. We presented results of two surveys on the reliability practices of public transport authorities. One of the main conclusions is that, most of times, service variability of vehicle performance is measured and monitored and a focus on passenger impacts is lacking. Second, there is no consistency in the definition of service reliability. We demonstrated that this may lead to different levels of quality concerning these indicators, while actual quality is constant.

We recommend taking passenger interest more explicitly into account when setting indicators and objectives of service reliability. In short headway services, regularity makes more sense than punctuality. Taking the actual number of passengers into account when aggregating the scores is also recommended. We demonstrated that an alternative indicator, being additional travel time (Van Oort & Van Nes, 2009b), represents the level of service reliability in a sustainable way, since it considers factors that are neglected by traditional indicators, such as driving too early and passenger boarding patterns.

The additional travel time was calculated for the tram lines in The Hague and compared to the indicators found in the international survey (presented in the previous section). It was demonstrated that no consistent result is possible, since different kinds of indicators are used. The location of measuring that differs between cities proved to be important too. This inconsistency may lead to wrong conclusions. The indicator of additional travel time incorporates the mentioned factors, enabling a more complete and consistent quantification of service reliability. Finally, if we are able to really monitor and analyse the passenger oriented indicators, we will find potential improvement measures and enhanced service quality will be achieved.

Although the survey results are very helpful to gain insights into how public transport authorities deal with service reliability, it is recommended to have an update. Such an update will illustrate the latest developments in this field and will also enable to extend the geographic spread of the participants (including Australia, Asia and South America for instance).

Acknowledgements

This research is performed in cooperation with KpVV, Dutch Knowledge Centre on Transport. We would like to express our gratitude to all participants of both surveys for their kind and much appreciated cooperation.

References

Bush, R. (2007). Does every trip need to be on time? Multimodal scheduling performance parameters with an application to Amtrak service in North Carolina. In *Proceedings of 86th annual meeting of Transportation Research Board*. Washington, D.C.

Ceder, A. (2007). *Public transit planning and operation, theory, modelling and practice*. Haifa: Technion-Israel Institute of Technology.

Chang, J., Collura, J., Dion, F., & Rakha, H. (2003). Evaluation of service reliability impacts of traffic signal priority strategies for bus transit. *Transportation Research Record, 1841*, 23–31.

European Commission. (2011). *White paper on transport*.

Frumin, M., Uniman, D., Wilson, N. H. M., Mishalani, R., & Attanucci, J. (2009). Service quality measurement in urban rail networks with data from automated fare collection systems. In *Proceedings of CASPT conference*. Hong Kong.

Hakkesteegt, P., & Muller, Th. H. J. (1981). Research increasing regularity. *Verkeerskundige werkdagen* (pp. 415–436) (in Dutch).

Hansen, I. A. (1999). *Report VK4810: Transport operation and management*. TU Delft.

Hickman, M. (2004). Evaluating the benefits of bus automatic vehicle location (AVL) systems. In D. Levinson, & D. Gillen (Eds.), *Assessing the benefits and costs of intelligent transportation systems*. Boston: Kluwer (Chapter 5).

Kimpel, T. J., Strathman, J., Bertini, R. L., & Callas, S. (2005). Analysis of transit signal priority using archived TriMet bus dispatch system data. *Transportation Research Record, 1925*, 156–166.

Landex, A., & Kaas, A. H. (2009). Examination of operation quality for high-frequent railway operation. In *Proceedings of 3rd international seminar on railway operations modelling and analysis*. Zürich.

Landex, A., & Nielsen, O. A. (2006). Simulation of disturbances and modelling of expected train passenger delays. *Computers in railways X* (p. 521). Southampton: WIT Press.

Please cite this article in press as: van Oort, N., Incorporating service reliability in public transport design and performance requirements: International survey results and recommendations, *Research in Transportation Economics* (2014), http://dx.doi.org/10.1016/j.retrec.2014.09.036

Levinson, H. S. (2005). The reliability of transit services: a historical perspective. *Journal of Urban Technology, 12*, 99–118.

Li, Z., Hensher, D. A., & Rose, J. M. (2010). Willingness to pay for travel time reliability in passenger transport: a review and some new empirical evidence. *Transportation Research Part E, 46*, 384–403.

Mazloumi, E., Curry, G., & Majid, S. (2008). Assessing measures of transit travel time variability and reliability using AVL data. In *Proceedings of 87th annual meeting of Transportation Research Board*. Washington, D.C.

Muller, Th. H. J., & Furth, P. G. (2000). Integrating bus service planning with analysis, operational control and performance monitoring. In *ITS 10th conference proceedings*. Washington, D.C.

Nakanishi, Y. J. (1997). Bus performance indicators, on-time performance and service regularity. *Transportation Research Record, 1571*, 3–13.

OECD/International transport forum. (2010). *Improving reliability on surface networks*.

Osuna, E. E., & Newell, G. F. (1972). Control strategies for an idealized public transport system. *Transportation Science, 6*(1), 52–72.

Pelletier, M., Trepanier, M., & Morency, C. (2011). Smart card data use in public transit: a literature review. *Transportation Research Part C: Emerging Technologies, 19*(4), 557–568.

Rietveld, P., Bruinsma, F. R., & van Vuuren, D. J. (2001). Coping with unreliability in public transport chains: a case study for Netherlands. *Transportation Research, 35A*, 539–559.

Schittenhelm, B., & Landex, A. (2009). Quantitative methods to evaluate timetable attractiveness. In *Proceedings of 3rd international seminar on railway operations modelling and analysis*. Zürich.

Schmöcker, J. D., Fonzone, A., Shimamoto, H., Kurauchi, F., & Bell, M. G. H. (2011). Frequency-based transit assignment considering seat capacities. *Transportation Research Part B, 45*, 392–408.

Shalaby, A. S. (1999). Simulating the performance impacts of bus lanes and supporting measures. *Journal of Transportation Engineering, 125*(5), 390–397.

Transportation Research Board. (2003). *Transit capacity and quality of service manual* (2nd ed.) Transit Cooperative Research Program, project report 100, Washington, D.C.

Van Oort, N. (2011). *Service reliability and urban public transport design*. T2011/2, TRAIL. (PhD thesis series). Delft http://www.goudappel.nl/media/files/uploads/2011_Proefschrift_Niels_van_Oort.pdf.

Van Oort, N., Boterman, J. W., & Van Nes, R. (2012). The impact of scheduling on service reliability: trip-time determination and holding points in long-headway services. *Public Transport, 4*(1), 39–56.

Van Oort, N., Sparing, D., & Brands, T. (2013). Optimizing public transport planning and operations using automatic vehicle location data: the Dutch example. In *MT-ITS conference*. Dresden.

Van Oort, N., & Van Nes, R. (2009a). Line length versus operational reliability: network design dilemma in urban public transportation. *Transportation Research Record, 2112*, 104–110.

Van Oort, N., & Van Nes, R. (2009b). Regularity analysis for optimizing urban transit network design. *Public Transport, 1*(2), 155–168.

Van Oort, N., & Van Nes, R. (2010). Impact of rail terminal design on transit service reliability. *Transportation Research Record, 2146*, 109–118.

Van Oort, N., Wilson, N. H. M., & Van Nes, R. (2010). Reliability improvement in short headway transit services. *Transportation Research Record, 2143*, 67–76.

Welding, P. I. (1957). The instability of a close interval service. *Operational Research Quarterly, 8*(3), 133–148.

Wong, C., & Wilson, N. H. M. (2006). Operations control strategies to improve transfers between high-frequency urban rail lines. In *Conference on computer aided scheduling on public transport*. Leeds.

Please cite this article in press as: van Oort, N., Incorporating service reliability in public transport design and performance requirements: International survey results and recommendations, *Research in Transportation Economics* (2014), http://dx.doi.org/10.1016/j.retrec.2014.09.036