Inventory and rating of performance indicators and organisational features in metropolitan public transport

A worldwide Delphi survey

Hirschhorn, Fabio; Veeneman, Wijnand; van de Velde, Didier

DOI
10.1016/j.retrec.2018.02.003

Publication date
2018

Document Version
Final published version

Published in
Research in Transportation Economics

Research in Transportation Economics 69 (2018) 144–156

Contents lists available at ScienceDirect

Research in Transportation Economics

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

Inventory and rating of performance indicators and organisational features in metropolitan public transport: A worldwide Delphi survey

Fabio Hirschhorn *, Wijnand Veeneman, Didier van de Velde

Faculty of Technology, Policy and Management, Delft University of Technology, The Netherlands

ARTICLE INFO

JEL classification:
L91
L98
R40
R48

Keywords:
Public transport
Performance
Organisational form
Delphi method

ABSTRACT

This paper contributes to the discussion on the ways organisational form of public transport systems may influence its performance. A worldwide Delphi survey with experts in the field is presented. After a three-stage iterative process interspersing questionnaires and controlled feedback the survey produces authoritative inventories and ratings of performance indicators and organisational features in public transport. In relation to performance indicators, system-wide metrics such as user satisfaction, cost-recovery, and modal split are selected as preferred measures for a strategic assessment of public transport. Concerning features of organisational form driving performance outcomes, integration emerged as the central dimension: policy integration between public transport and other sectors, single integrated planning authority, as well as ticket and fare integration were highly rated by Delphi experts.

1. Introduction

The connection between organisational form and performance in metropolitan public transport ("PT"),[1] has been under analysis for over a century - Chadwick (1859) illustrates this - and remains at the top of the academic agenda (Docherty, Shaw, & Gather, 2004; Faivre d'Arcier, 2014; Hensher & Wallis, 2005). These analyses indicate that some organisational features (market deregulation, competitive tendering etc.) might be important for achieving goals attached to PT (such as efficiency, accessibility etc.). However, unravelling this relationship is not simple and studies many times have mixed or inconclusive results (Roy & Yvrande-Billon, 2007; van de Velde & Wallis, 2013). Reasons for these difficulties are manifold. The first challenge is the identification of both performance and organisational variables to be examined.

Concerning **performance**, two levels of analysis exist: the first refers to broad PT aims (e.g. efficiency and accessibility). Complexities emerge as the definition and prioritisation of broad aims change over time and according to the varied, and sometimes conflicting, expectations actors have in relation to PT. The second level of analysis is related to the translation of those broad aims into quantitative metrics - performance indicators (for example vehicle hour per employee, population living within 500m to frequent PT service). The choice of the best indicator or set of indicators to measure broad aims varies substantially and consensus is a challenge (Fielding, 1992). Features of **organisational form** normally have an eminently contextual character. As a result, their study requires in-depth case knowledge and the generalisation of conclusions about their influence over performance is arduous.

In view of this complexity, most studies connecting organisation and performance tend to assess the impacts of one single policy creating or reforming a feature of PT organisation on one type of performance. For instance, the potential cost impacts stemming from the use of competitive tendering in bus markets (Beck, 2011). When framed this way, these analyses tend to simplify reality by only examining a direct and simple relationship between isolated organisational features and performance indicators. A more comprehensive approach looking at combined effect of multiple variables and their dependencies might have the potential to unravel new insights on key mechanisms connecting organisation and performance in PT.

A first step, therefore, is to identify variables that might enable an analysis of PT not driven by the assessment of a single policy, but rather one that considers the interplay of different elements. For this purpose, this article aims at defining *what performance metrics are suitable to*

* Corresponding author.
E-mail addresses: f.hirschhorn@tudelft.nl (F. Hirschhorn), w.w.veeneman@tudelft.nl (W. Veeneman), D.M.vandeVelde@tudelft.nl (D. van de Velde).

[1] Metropolitan public transport, abbreviated to PT, refers to all collective modes of land passenger transport services available to the general public within a metropolitan area and linking it to its direct environment. There is no distinction based on ownership or control - these services could be either publicly or privately operated.

https://doi.org/10.1016/j.retrec.2018.02.003
Received 30 November 2017; Received in revised form 19 February 2018; Accepted 23 February 2018
Available online 11 March 2018

measure strategic PT outcomes and what organisational features might drive strategic outcomes.

The Delphi method (Dalkey & Helmer, 1963; Linstone & Turoff, 1975) is a possible tool to help with this task. The methodology consists in a participatory process to elicit and articulate the opinion of experts in a field, allowing them to expose diverse views on complex matters. Whilst the method was initially devised as a tool for consensus reaching, variants emerged enabling researchers to produce different sorts of outcomes. The Delphi method can have an important role in highlighting multiple perspectives, clarifying policy problems and assisting problem-resolution. In this article, a Global Delphi in Public Transport ("GDPT") gathering PT experts across the world is presented. After an iterative three-stage process, the GDPT produced authoritative inventories and ratings of core performance indicators and organisational features driving performance in PT.

The article starts by reviewing literature on PT performance and organisational form, to then outline the suggested alternative analytical approach (2). In continuation, the GDPT is described (3). Outcomes are discussed (4), followed by conclusions (5).

2. Background

2.1. Performance

To disentangle the topic of PT performance two levels of analysis can be distinguished, one involving performance aims - associated to an *ex-ante* focus on policy design and interventions – and another related to performance indicators – that have an *ex-post* focus on policy outcomes.

The first level is tied to public values. Public values can be described as general abstract principles defining government responsibilities and rights, and obligations of citizens (Jørgensen & Bozeman, 2007). More concretely, public values represent the varying goals and expectations of stakeholders in relation to government in a policy area (de Bruijn & Dicke, 2006). As such, values are constrained by the surrounding institutional setting, and their definition and prioritisation vary in context. Furthermore, values may conflict and compete with each other (Thacher & Rein, 2004). In PT, values such as efficiency, accessibility, safety, etc. are at play. The importance attributed to each of them may change in time, according to actors' interests, and also depending on context. This is reflected in literature that describes paradigm shifts in PT. Banister (2008) and Marshall (2001) identify the emergence of a 'sustainable mobility paradigm' replacing a conventional approach to transport planning ('neoclassic' or 'predict and provide'). The latter emphasises efficiency and utility - mobility is seen as a derived demand. The 'sustainable mobility paradigm', differently, acknowledges social and environmental perspectives too, and encourages sustainable transport patterns. Similarly, authors identify the rise of the accessibility paradigm supplementing a mobility-centred view, giving more prominence to issues of spatial and social inequality (Farrington, 2007).

Additionally, as public values, broad PT aims may also compete with each other. Buehler & Pucher exemplify this: "In general, more frequent, higher quality, and financially efficient public transport also helps achieve social sustainability. However, in some circumstances financial efficiency and social equity might not be fully compatible (BBR, 2002; Topp, 2006; Werner, 2006)." (2011, p. 135). Walker distinguishes between patronage goals - those achieved by the extent people use PT - and coverage goals - concerned with availability of PT regardless of its use, and asserts: "Public transport must serve the competing demands of patronage and coverage, because the two values push service design in opposite directions" (2008, p. 442).

The second level of analysis of performance relates to the quantitative measures through which broad aims can be translated – the performance indicators. The choice of indicator or set of indicators to measure each performance aim varies and consensus is a challenge (Fielding, 1992). Firstly, the range of options is vast: Geerlings, Klementschitz, and Mulley (2006) report literature review identifying over 400 indicators in PT. Furthermore, choices may vary due to particular objectives of the analysis undertaken, type and amount of data available, or methodology employed. Taking efficiency as an example: Veeneman (2002) justifies his choice to measure efficiency with cost-recovery ratios based on his interest in a metric for the whole public transport organisation and from a general policy perspective. Fielding (1992), proposing efficiency metrics to be adopted by PT agencies in the USA, suggests 5 indicators - revenue vehicle hours per dollar of operating expense; vehicle miles per peak vehicle; vehicle hours per employee; vehicle miles per maintenance employee; vehicle miles per accident). Jain, Cullinane, and Cullinane (2008) employ Data Envelopment Analysis to compare technical efficiency in 15 PT systems measuring the number of vehicle kilometres and passenger trips (supply and demand-oriented metrics).

2.2. Organisational form

Different authors describe the organisational structure of PT systems (e.g. Barter, 2008; van de Velde, 1999). Based on these models, it is possible to identify some important organisational features: existence of a governmental prerogative to set-up public transport services or to authorise others to do so (as opposed to deregulated regimes with autonomous market entry), division of regulatory powers between government tiers, ownership structure of operating companies, varied awarding mechanisms, etc. The individual impact of these and other features over performance is scrutinised in literature.

The contrast between markets with open entry to autonomous players and markets where the public sector holds a 'legal monopoly' to initiate PT services constitutes an important research track in PT. Authors examine both deregulation experiences (Cowie, 2014; Paredes-Molina & Baytelman, 1996) as well as systems adopting regulated entry (Gómez-Lobo, 2007; Zhang, Juan, & Xiao, 2015). Particular elements of 'regulated' markets are also evaluated: different awarding mechanisms or the varied contractual regimes (risk allocation and incentives) governing the relationship between government authorities and operating companies (Kavanagh, 2016; Stanley & Hensher, 2008).

Authors also examine how ownership issues affect PT performance, both at the market level – integration or separation between infrastructure management and service delivery (van de Velde et al., 2012) - and at operating company level (Albalate, Bel, & Calzada, 2012).

Analyses also cover how having different tiers of government responsible for PT may influence performance. Likewise, the role of different key PT stakeholders, and the ways they interact is studied (Buehler & Pucher, 2011; Finn & Mulley, 2011). Finally, the performance implications of funding frameworks (Veeneman, Augustin, & Enoch, 2015), and of service characteristics (Brown & Thompson, 2008) are also analysed in PT literature.

2.3. A proposed approach

PT is a complex multifarious socio-technical system where technical elements and actors with diverse and conflicting values coexist. Whilst the analyses described in Sections 2.1 and 2.2 offer important insights about the connection between organisational form and performance, they do so by examining variables in an isolated manner. Thus, they might be unable to capture a more nuanced view of the complexities of PT systems. Literature may be overlooking relevant mechanisms linking organisation and performance. This echoes similar claims: Hale (2011) searches for new approaches to assess urban transport, as "Most transport assessment is generally based on incremental analysis of individual projects with pre-existing planning and political support." (p. 173).

It is important to acknowledge the complexity in PT systems and search for ways that, even if only gradually, allow addressing the relationship organisation-performance in a more comprehensive way, rather than looking at policy processes as the summation of isolated interventions. The connection between organisational form and performance can be better understood from a configurational perspective.

Variables interact affecting and being affected by each other, and thus conjunctural causation processes can occur (Ostrom, 2010; Ragin, 1987). It is plausible to expect that PT performance outcomes result from the effects of a combination (configuration) of different organisational variables. Moreover, more than one combination of organisational variables may lead to the same outcome.

A necessary step in the search for new analytical approaches in PT is to find metrics that can help measuring strategically important transport initiatives or projects. Those metrics should be able to identify "which kind of options and interventions are more able to deliver substantial and hence strategically important improvements to overall urban transport outcomes." (Hale, 2011, p. 175). This article aims at defining *what performance metrics are suitable to measure strategic PT outcomes and what organisational features might drive strategic outcomes*. The Delphi method is employed for this.

3. Global Delphi in public transport

3.1. The Delphi method

The Delphi method consists in a participatory process for consensus building. It elicits the opinions of experts through a series of questionnaires interspersed with controlled feedback to build authoritative forecasts in relation to the occurrence of events or trends (Dalkey & Helmer, 1963). After responding to the first questionnaire, and preceding new rounds, participants have access to, and can reflect on, other experts' opinions through anonymous feedback organised by the survey's coordinator.

Variants of Delphi emerged highlighting objectives different from consensus. The Policy Delphi, 'seeks to generate the strongest possible opposing views on the potential resolutions of a major policy issue' (Linstone & Turoff, 1975, p. 84). The ranking-type Delphi is used to establish the relative importance of issues building authoritative rankings (Schmidt, 1997). This paper combines and adapts these two variants.

The Delphi can help articulate different views and estimations on complex matters, and also allows the confrontation of perspectives to generate ideas and shed light on alternative directions for clarifying and solving problems. It constitutes an important tool for answering complex issues that cannot rely on the knowledge of a single expert or single group of stakeholders. The anonymity of the process supports free expression of opinion and avoids that a more vocal individual or group control the discussion. Furthermore, the method circumvents the practical difficulty of bringing experts together at a same location at the same time. Finally, the method permits engaging stakeholders of multiple affiliations.

3.2. The GDPT

A Global Delphi in Public Transport ("GDPT") was structured in three different stages: (i) brainstorming (respondents could freely propose all relevant elements in connection to the issues at stake), (ii) narrowing-down (respondents shortlisted most relevant elements from previous stage) and (iii) rating (respondents rated shortlisted elements). Each of these stages used one online questionnaire. A dedicated blog was also created to support the survey as an additional communication channel: the blog served as a platform for the publication of results and updates, and to provide information on the survey's motivation and aims.

3.2.1. Choice of experts

The choice of experts is crucial to ensure breadth of knowledge amongst panellists (Delbecq, Van de Ven, & Gustafson, 1975). Panel-building in a Delphi has two moments: (i) defining the relevant expertise and (ii) identifying individuals with that knowledge.

Concerning relevant expertise, the GDPT's interest was knowledge on (i) PT performance monitoring and/or evaluation and (ii) PT governance - design and functioning of organisational forms adopted in diverse PT systems worldwide. For the identification of experts two common approaches are sampling based on actor types and snowball sampling. The first approach presumes that representativeness in terms of actor type's guarantees representativeness in terms of perspectives and aims to include stakeholders based on diversity of affiliation. In snowball sampling, the researcher starts off by picking a small number of stakeholders, and then asks them to mention other potential participants (Cuppen, 2010). The GDPT combined both approaches.

The sampling based on actor's type, followed some operationalisation steps:

a) First, the GDPT aimed at including individuals of prominence in the field, i.e. current or previous affiliation to eminent organisations, as well as active involvement in major international *fora*, major universities, government entities responsible for PT, PT providers, or participation in editorial boards of prominent international journals.
b) Second, the GDPT targeted a blend of stakeholders in all relevant roles in PT, ensuring the inclusion of multiple views, i.e. (i) academics, (ii) government officials, (iii) employees from transport operating companies; (iv) users' associations, (v) employees of multilateral institutions, (vi) consultants. In the case of academics, two more aspects were considered: works published in relevant international journals and retrieved on Google Scholar, and variety of views based on academic discipline: transport geography; transport economics; transport engineering; public administration and policy; and urban planning (Veeneman, 2002).
c) Finally, the survey sought experts based in and/or with expertise on varied geographical locations (ensuring the global character of the GDPT).

Based on these characteristics, a simple matrix was built to help the selection. A first attempt to populate the matrix found around 170 names. Some of the authors' professional contacts were also part of the list. At this point, a sample of the initially identified experts was contacted and asked to provide recommendations of other experts to participate in the survey (snowball sampling) – names suggested that had not been identified previously were also included in the matrix. All experts identified after these steps were assessed more closely to confirm the direct relevance of their work for the topics being surveyed - not every expert in PT is necessarily knowledgeable on the issues at hand.

Invitations to participate were sent to experts by email including the link to the first questionnaire. Based on the number of emails that bounced back and the number of accesses to the questionnaire, it is possible to confirm that 96 experts received the first link. The first questionnaire was fully responded by 54 experts. From those 54, 48 participants completed also the second questionnaire. A final group of 46 participants from 18 different countries[2] concluded the third and last questionnaire completing the entire survey. In a Delphi, the survey's coordinator has no ability to enforce participation of invited experts and having a low turn-out is a significant risk. However, the GDPT's numbers are very positive: response rate was high, drop-out along the survey low, and the profile of respondents shows the desired diversity, including approximately 60% practitioners and 40% academics, varied technical and regional knowledge (Fig. 1).

Statistics are based on self-stated information by respondents. Respondents could choose multiple options. Percentages are in relation to the total number of participants.

3.2.2. Brainstorming

The first questionnaire of the Delphi used both open-ended and Likert-scale questions.

In the open-ended questions, experts were asked to assume they

[2] Australia, Brazil, Canada, Chile, Colombia, Denmark, England, France, Italy, Japan, Netherlands, New Zealand, Norway, Singapore, South Africa, Sweden, Turkey, USA.

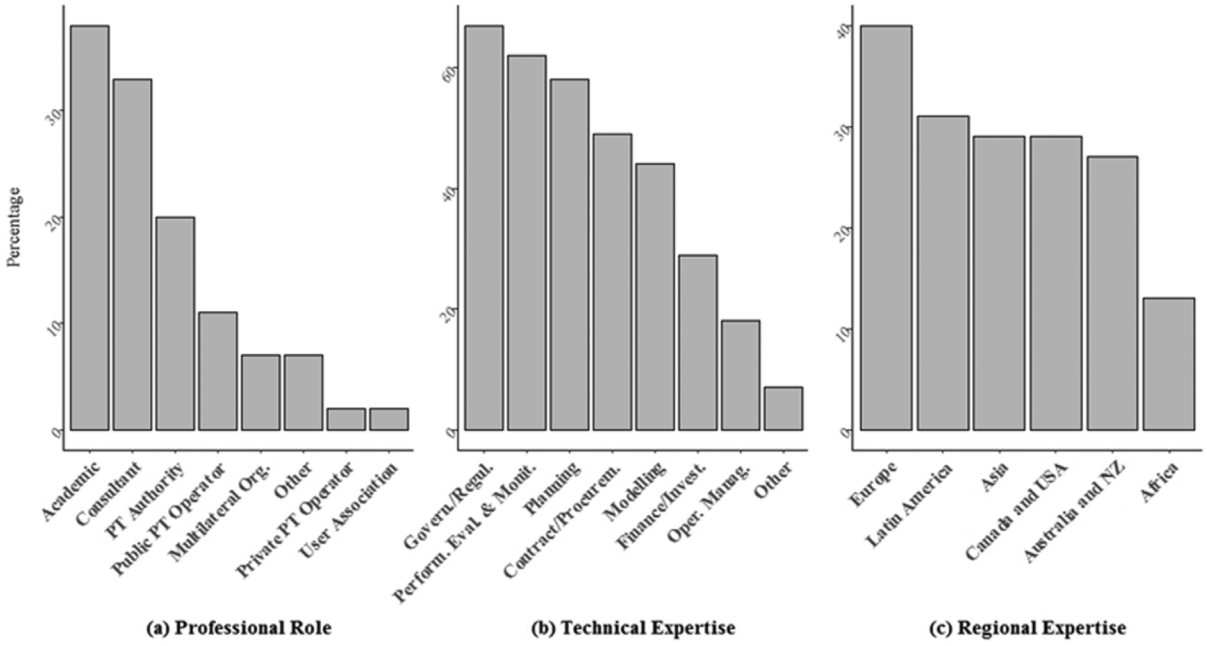

Fig. 1. Profile of experts that concluded the GDPT.

would be hiring a consultant to analyse a PT system and, thus, had to list and briefly describe at least 5 performance indicators that they would want to have examined by the consultant. Experts were instructed to prioritise indicators that, in their view, would be better able to provide relevant information on critical aspects of PT (there was no instruction as to what critical meant). By not imposing a limit to the number of indicators as well as not defining any specific performance dimension to be measured, experts were given free room to expose their views. In addition, experts were also asked to associate each indicator they had listed to a broad performance aim. This 'pairing' of indicator and aim followed the idea of two levels of analysis in performance described in 2.1 above and that had been described to experts in the questionnaire. A second similar question requested experts to list and describe at least 5 organisational features that, on their view, would be important drivers of PT performance. Again, there was no request for considering any specific type of performance. The question clarified what was meant by organisational feature - defined in the questionnaire, based on the literature described in Section 2.2.

In the following set of questions, experts were presented with 'pairs' of performance aims and indicators that are frequently discussed in PT literature (e.g. *Environmental Sustainability;* per capita *emissions of NOx)*. This question was deliberately asked after the open-ended questions, avoiding any bias. The 'pair' structure was again consistent with the idea of two level of analysis of performance used in the questionnaire. Experts had to rate the relevance of these 'performance pairs' in a scale from 1 to 5. The same rating type of question was proposed in relation to a literature-based list of PT organisational features (e.g. *Allocation of Ownership of Long-Life Assets)*. These rating questions were included in the GDPT as 'fall-back options': in case the open-ended questions had generated answers that would need to be discarded (a frequent problem in Delphi), the responses to the rating questions could be used to inform the following rounds of the survey reducing the loss of expert input. This safety-net was not needed though, as the open-ended questions generated nearly 700 lines of content fitting the research aim.

3.3. Analysis of responses

Answers to open-ended questions were qualitatively analysed. Major themes were identified, grouped, and redundancies eliminated. One author performed an initial revision for the entire set of answers. As a reliability test, the second author independently coded a randomly selected large sample of the answers related to performance, and the third author, also independently, coded a randomly selected large sample of the answers related to organisational form. The triangulation of these assessments afterwards showed that virtually all coding made by the authors was coherent. Minor differences were discussed and reconciled to generate the final output of this round.

Revising and coding such large amount of information was challenging and required striking a balance between, on one hand, the task of consolidation to produce a reasonably-sized list to be used in following rounds by experts (that have limited time available for participation) and, on the other hand, avoiding generalization - that would defeat the purpose of the GDPT. Having asked the experts for a brief description of their input was crucial to help clarifying opinions and enriching the material. One further important aspect of the qualitative coding was the use of sensitising concepts - general conceptual guidelines derived from existing literature and used to support the interpretation and organisation of input received from experts (Bowen, 2006).

In relation to **performance**, the evaluation frameworks developed by Dajani and Gilbert (1978) and Fielding, Babitsky, and Brenner (1985) were taken as source of sensitising concepts. These frameworks define three broad classes of performance measures (i) *cost-efficiency* (technical relationships between service input and service outputs); (ii) service *effectiveness* (the degree to which PT achieves mobility goals); (iii) *cost-effectiveness* (relationship between service inputs and consumption of services). In addition, they also include a fourth dimension, *impact measures*, that account for indirect, beneficial or negative, intended or unintended impacts of PT on social well-being, economic development, and environment.

These frameworks were not taken prescriptively, and only served as guidelines for the interpretation and organisation of answers received. One example of how the actual responses shaped the brainstorming output involved the attempt to have experts associate broad performance aims and indicators – the 'pairing' exercise. During the design of the first questionnaire, the possibility that this would raise difficulties was foreseen. Firstly, because terminology in PT may be very confusing, especially when experts from different geographical regions dialogue. There were cases in which same indicators were associated to different broad aims by respondents, highlighting how difficult it is to conceptualise and operationalise these broad values. Language and professional jargon

need particular care in Delphi surveys to avoid misinterpretations. Furthermore, answers showed that the conceptual discussion on broad performance aims was mostly restricted to academics - practitioners in the panel hardly used definitions as efficiency, effectiveness etc. As a result, the authors decided to not persist with the discussion of broad aims and restrict the performance discussion in following rounds of the GDPT to performance indicators, lest participants could lose interest in the survey and no meaningful result would be achieved.

The coding process identified the following broad performance themes emerging from experts' answers: 'Supply', 'Cost and Revenue', 'Financial Sustainability', 'Usage', 'Service Quality and User Satisfaction', 'Accessibility Impacts', 'Wider Impacts', 'Other'. These thematic blocks were divided in 38 clusters that contained a total of 109 performance indicators. An inventory of performance indicators was built (Annex 1).

Concerning **Organisational Features**, van de Velde (1999) was used as a source of sensitising concepts. The author describes the varied forms PT organisation assumes depending on actors involved, the way their relationship is governed, and the allocation of strategic, tactical and operational tasks amongst these actors. These concepts were not employed prescriptively and responses provided by experts shaped the coding process.

Answers revealed again how convoluted the use of terminology in PT may be. The term 'regulator', for instance, was many times used to designate the body responsible for the integrated planning of PT, but was also used to refer to the entity responsible for enforcing regulations in the industry (watchdog tasks). These roles are sometimes, but not always, combined in one entity and the use of the same denomination is a common cause of misunderstandings. Additionally, many participants went beyond the consideration of formal structural elements of PT and also listed 'soft' elements, such as the skill-set of staff, trust, and leadership for instance.

After the coding, the following major themes were identified: 'Initiative and Funding', 'Planning', 'Operations', 'Contracting Practices', 'Control and Accountability', 'Other Integration and Fragmentation Features, 'People and Relationships', 'Other'. These thematic blocks were divided in 24 clusters containing 70 organisational features. A final inventory of PT organisational features was produced (Annex I).

The feedback material sent to experts after the Brainstorming round included the two full inventories (performance indicators and organisational features) and comments respondents had included in their answers.

3.3.1. Narrowing-down

The second round of the GDPT consisted in a shortlisting exercise. Experts were asked to select 7 performance indicators and 7 organisational features amongst all those in the inventories produced in Round 1. The question requested experts to prioritise indicators better able to provide insights on most critical aspects of PT and organisational features with greater impact on performance. Proposing a limit of 7 items per shortlist envisaged emphasising the need for prioritisation, given the very large inventories that had been built in Round 1[3]. It was also a way to limit the workload of respondents, a pre-emptive measure to reduce possible drop-out.

3.4. Analysis of responses

Two types of analysis were performed in relation to the responses received from experts. A first approach considered votes given at the level of individual variables, i.e. a simple count of votes that each performance indicator and each organisational feature received. A second examination was done for cross-checking results. Responses were also examined at the cluster level, i.e. the analysis of responses took into account what would have been the sum of votes given to each of the clusters that had been defined in the coding of Round 1. For instance, the performance cluster 'Total (and operating) costs ratios' defined in Round 1 comprised nine different indicators, so all votes given to these nine indicators were added to assess the total votes of the respective cluster. This second analysis revealed that selecting a shortlist of 7 items for the subsequent Round of the GDPT would not adequately reflect experts' priorities. The cluster 'Total (and operating) costs ratios' is again an example – it was the third most voted cluster in the inventory of performance indicators, but since these votes were dispersed amongst the nine alternative individual indicators, none of these nine would, individually, be in the final shortlist of 7. Therefore, whilst clearly relevant to experts, no cost ratio would have made the cut. Authors decided to increase the size of the shortlists to 10 variables each (results in Table 1).

In the comments accompanying their responses in this Round, some experts manifested that the question proposed in the GDPT should have defined a specific policy objective to serve as a guide for the shortlisting of performance indicators, e.g.: '*All performance is about your objectives; if you clarify that, you answer the question asked. Objectives vary between systems and over time*'. The GDPT deliberately avoided instructing experts to attach their choices to the assessment of specific policies or performance aims. This was coherent with the objective of using the GDPT as a first step for a more comprehensive approach to PT (Section 2.3). The fact that some of the experts were expecting to receive a specific policy or goal on which to base their opinion suggests that the premise of the article is accurate and that the GDPT managed to frame questions in accordance to its purpose.

The two final shortlists, presented in Table 1, were defined by majority of votes (Schmidt, 1997). Four experts voted for more options than requested in the questionnaire so their choices were considered based on weighted values so that results would not be skewed. The feedback material sent to respondents included the two shortlists of 10 items, the ratio of votes each variable had received, and also other comments made

Table 1
Summary of results Round 2.

Performance Indicators	Respondents (%)
Cost-recovery ratio	45%
Modal split: by trips and passenger km	41%
User satisfaction (overall index)	37%
On-time performance according to timetable	31%
Ratio between travel time in PT and car	30%
% of inhabitants (or users) living within walking distance to frequent PT service	30%
Ridership per capita	27%
Cost per passenger km	21%
Total revenue and total cost	17%
Comfort: average vehicle occupation, fleet conditions and characteristics	16%
Organisational Feature	**Respondents (%)**
Integrated fare and ticketing (clearinghouse functions)	59%
Funding framework: source, availability, autonomy/control	57%
Integrated planning: multimodal and multijurisdictional agency or capability	43%
Policy Integration: cross-sector links with other govt. areas	39%
Clear legal and regulatory frameworks: mandates and performance targets	34%
Long-term strategic PT plan	31%
Contracts: risk allocation and incentive structure	28%
Skill set and technical expertise of staff	27%
Awarding mechanism employed	23%
Level of competition among operating companies (re. market concentration)	21%

These are summarised descriptions of variables and not the full descriptions defined in the first Round (included in Annex I).

[3] The specific number of 7 performance indicators and 7 organisational features was suggested because this was the average number of performance indicators listed by each respondent in Round 1.

by experts.

3.4.1. Rating

In the final questionnaire experts were asked to allocate a total of 100 points amongst (all or part of) the 10 shortlisted performance indicators and to do the same for the shortlisted organisational features. Points were to be freely distributed to reflect relevance of variables if experts had to use them for a comparative study of PT in 15 metropolitan areas. The stated objective of this hypothetical comparative study was to unveil how different ways of organizing PT influence different types of performance.

Wording in the questionnaire once more referred to the performance indicators' ability to provide insights on strategic aspects of PT and to the organisational features' ability to drive performance. Furthermore, to avoid inducing bias, variables were presented in a random fashion in each individual questionnaire, so each participant saw a list ordered differently.

3.5. Analysis of responses

Answers were compiled and measured in different ways (described in Table 2): (i) the average points received by each variable ('Avg. Points'), (ii) the standard deviation of points received ('Std. Dev.'), (iii) the highest single score attributed to the variable by one expert ('High'), (iv) the percentage of experts attributing zero point to a variable ('Zeros'), and (v) the rank of variables based on the amount of points they received ('Rank'). A comparison with their ranking in the previous round is also possible ('Rank 2').

4. Results and discussion

4.1. Results

Overall choices made by experts reveal some central elements in the discussion organisation-performance in PT. In relation to performance, answers manifest a preference for a high-level system-wide assessment of PT using multipart indicators. Concerning PT organisation, integration emerged as the central dimension: policy integration, integrated strategic planning and integration of tasks and system elements.

On a more micro perspective, the GDPT's results provide a 'menu' of core performance indicators and organisational features. These twenty elements (ten and ten respectively) spell out the broader views highlighted just above. A first catalogue of remarks in relation to this 'menu' is proposed, nonetheless it warrants further investigation. Importantly, these remarks look at results from a particular perspective, i.e. they consider the article's purpose to define *what performance metrics are suitable to measure strategic PT outcomes and what organisational features might drive strategic outcomes*.

4.1.1. Performance indicators

Interactions and overlaps. The analysis of **performance indicators** shortlisted in **Round 2** shows interactions and/or overlaps within the ten most voted metrics, e.g. '*User Satisfaction*' and '*Comfort*' (the latter is a component of the first). The same with '*Total Revenue and Total Cost*' and '*Cost per Passenger Km*'. Rather than a problem with the survey, these 'redundancies' demonstrate the consensus amongst experts around the importance of certain broad aims – e.g. the pervasiveness of PT can be associated to at least six indicators in the list. At the same time, this underscores the controversies on the choice of best indicators to translate broad aims (as mentioned in Section 2.1).

Patronage goals. Another aspect of consensus visible in the shortlist is the prevalence of indicators associated to patronage goals (as defined by Walker (2008)). The only indicator directly related to coverage goals is '*% of inhabitants (or users) living within walking distance to frequent PT service*'.

'Big Picture' system-wide indicators. The three most voted indicators ('*Cost-recovery Ratio*', '*Modal Split*' and '*User Satisfaction*') are system-wide measures composed by other variables. They provide a 'big picture' of PT systems rather than information on specific service elements.

Noted absences. Amongst absences from the shortlist, '*Vehicle Km per Capita*' and '*Subsidy Level*' had voting rates close to the 'top 10' items. Furthermore, none of the environmental impact or affordability indicators included in the first round's inventory was shortlisted. The 'best ranked' amongst these two categories was '*Percentage of income or household budget spent on transport*', at only the 26th overall position. This is surprising considering the increasing attention environmental and affordability aspects receive currently.

Results from **Round 3** add new angles to these observations. When faced with the task of comparing the relative importance of shortlisted indicators, some participants realised interactions and overlaps that had resulted from Round 2. One participant mentioned, for example: '*Some of these [performance indicators] are substitutes or inputs to others,* e.g. *comfort is an input to customer satisfaction arguably leading to a need only for the latter.*' These observations corroborate the point discussed above.

'Big Picture' system-wide indicators again. The three most voted indicators remained the same between Rounds. The upshot in Round 3 is the visible gap in points separating them from the rest of the list. This

Table 2
Summary of results Round 3.

Performance Indicators	Avg. Points	Std. Dev.	High	Mode	Zeros	Rank	Rank 2
User Satisfaction (overall index)	15,91	11,03	50	15	11%	1	3
Cost-Recovery Ratio	15,24	9,07	30	20	14%	2	1
Modal Split	13,20	9,68	40	20	20%	3	2
% of Inhabitants (or users) living within walking distance to Frequent PT Service	9,78	7,71	30	10	23%	4	6
Ridership per Capita	9,57	8,70	30	5	25%	5	7
Ratio between Travel time in PT and Car	8,22	6,36	25	10	25%	6	5
On-time Performance according to Timetable	8,04	6,95	20	0	32%	7	4
Total revenue and total cost	7,28	8,39	30	0	45%	8	9
Cost per Passenger Km	6,98	6,77	26	0	36%	9	8
Comfort	5,78	5,88	20	0	41%	10	10
Organisational Features	**Avg. Points**	**Std. Dev.**	**High**	**Mode**	**Zeros**	**Rank**	**Rank 2**
Policy Integration: cross-sector links with other govt. areas	14,65	7,79	30	10	9%	1	4
Funding: source, availability, autonomy/control	14,37	11,20	50	15	18%	2	2
Long-term Strategic PT plan	12,65	7,82	30	10	11%	3	6
Integrated Planning: multimodal and multijurisdictional agency or capability	12,46	8,42	40	10	16%	4	3
Integrated Fare and Ticketing (clearinghouse functions)	10,28	8,81	40	10	25%	5	1
Clear Legal and Regulatory Frameworks	8,96	6,87	30	10	25%	6	5
Contracts: risk allocation and incentive structure	8,43	7,50	30	10	27%	7	7
Skill set and technical expertise of staff	7,43	6,96	25	0	34%	8	8
Awarding Mechanism Employed	4,50	5,09	20	0	48%	9	9
Competition among Operating Companies (re. market concentration/fragmentation)	4,09	4,55	15	0	50%	10	10

underscores the preference that consulted experts manifested in relation to multipart measures that provide information on ample aspects of PT.

A controversy with 'User Satisfaction'. Whilst the metric became the highest rated amongst all shortlisted indicators (15,91 points on average), expert's views on its importance proved to be also highly dispersed (standard deviation above 11), denoting fragmented opinions.

Consensus. Opinions around *'Comfort'* are less fragmented. This indicator had the lowest standard deviation (5,88) and a high percentage of 'zeros' (41%), both results suggesting lower relative importance. This may be caused by the overlap with *'User Satisfaction'* as highlighted in the comment from the respondent included above. The same reasoning may explain the variation in ranking of the indicator *'On-time Performance according to Timetable'*, i.e. it lost positions possibly for being a component of user satisfaction.

Lower informative value of 'Total Revenues or Total Costs. The very high percentage of 'zeros' attributed to 'Total Revenues and Total Costs' (45%) corroborates opinions expressed by some experts advocating that cost measures are more informative per unit of production. The presence of this indicator in the shortlist might be a reflection of terminology misinterpretations during the Delphi.

4.1.2. Organisational features

Consolidation vs fragmentation debate. The analysis of the shortlist of **organisational features** produced in **Round 2** indicates that, according to consulted experts, the way tasks are allocated amongst actors or the way system features are combined or separated represent the core organisational dimension driving performance. The debate on consolidation and fragmentation - either of tasks and roles amongst PT actors, or in relation to PT system features – appears in at least four features amongst the top ten - *'Integrated Fare and Ticketing'*, *'Integrated Planning'*, *'Policy Integration (cross-sector links with other govt. areas)'*, *'Long-term strategic PT plan'*. The selection of all these elements reinforces the importance attributed to coordination as a central aspect in PT.

Integration as Coordination. Comments from experts indicate that the rationale associating integration and coordination seems to have guided some of their opinions, for instance: *'A unique organisation responsible for the planning of the transport issues of the city as a whole, considering all modes and making long term plans.'*. However, evidence suggests that formal integration through hierarchic governance is not a condition for coordination, either in PT (Chisholm, 1992) or more broadly (Ostrom, 1990).

Consensus. The shortlisting exercise promoted in Round 2 shows a visible consensus amongst experts in relation to the importance of the two most voted features. *'Integrated Fare and Ticketing'* and *'Funding: source, availability, autonomy/control'* had a clear vote advantage in relation to the remaining features.

Interactions and overlaps. As with performance indicators, the shortlist of organisational features has a series of components that interact and/or overlap. As such, a clear emphasis on Strategic and Tactical tasks (as defined by van de Velde (1999)), particularly PT policy design and planning, is evident.

'Old favourites' in the bottom of the list. Liberal reforms taking place in the last decades sought to reduce public spending in PT. Regulatory changes and business practices targeted rules for market access and competition as crucial mechanisms to promote these efficiencies. As such, 'Awarding *Mechanism Employed'* and *'Competition amongst Operating Companies'* were prominent features in discussions related to PT, seen as strong performance drivers (as described in Section 2.2). The GDPT shows a different scenario. It is true that these two features are shortlisted, however they hardly made the cut. Experts seem to have concluded that these features are less impacting than previously imagined.

Noted absences. *'Regulatory agency or capability (watchdog)'* and *'Business structure of operating companies (formal versus informal paratransit models)'*, included in the inventory produced in Round 1, have been also at the centre of recent policy and academic debates. Nonetheless, they are not shortlisted.

Other interesting aspects involving organisational features emerge in **Round 3.**

Coordination at the core - consolidation vs fragmentation debate strengthened. Consistently with Round 2, PT features of integration (of tasks or system characteristics) are emphasised by experts. These features are four of the five with highest average score, and a low percentage of experts attributed a 'zero score' to any of them.

Interactions and overlaps. The position of PT policy design and planning as important drivers of performance is strengthened when average points are considered. Interactions and overlaps may have contributed to this outcome, as seen from some comments provided by experts: *'My inclusion of 'Long-term Strategic Plan' assumes that the plan was developed with robust engagement of local and state agencies that provide policy and* funding *guidance as well as agencies and functions that have an impact on (leverage) transit operations - land use, economic development, housing, etc.'*.

Dissention in relation to funding. Round 3 shows mixed results in relation to the importance attributed by experts to funding practices: whilst this feature received the second highest amount of points, it also shows the highest standard deviation amongst all features. This repeats the situation observed above with the performance indicator *'User Satisfaction'*. It may be a result of how questionnaire three was framed, proposing a selection of features for a comparative study - maybe experts consider fare and ticketing integration less informative if used for this purpose.

A confirmed consensus. There is again some consensus on the relative lack of relevance of 'Awarding *Mechanism Employed'* and *'Competition amongst Operating Companies'*. Both remained at the bottom of the ranking after votes in Round 3. Not only they received less points, but also opinions in relation to these two features are less dispersed (they show the lowest standard deviation rates) and they received 'zeros' from 48% to 50% of respondents respectively.

Ranking changes. *'Policy Integration (cross-sector links with other govt. areas)'* topped the ranking in terms of points received, whilst *'Integrated Fare and Ticketing'*, that had the highest percentage of votes in Round 2 moved to the fifth position. One possible explanation for this is the framing of the question in Round 3, as speculated for the dissention in relation to funding.

4.2. Experts and method

The GDPT combined elements of a Policy Delphi (evoking multiple conflicting views on issues) and a ranking-type Delphi (building authoritative ratings). It was effective in gathering a diverse set of experts and promoted a qualitative exercise - by design not statistically significant - with results that could hardly be achieved through a different way. The periodic feedback helped in keeping experts engaged - the high response rates testify to this. In addition to the feedback reports, a dedicated blog was created and used to post survey's details and updates. Finally, by introducing the point allocation methodology instead of a simple ranking question in the last round, the GDPT does not only measure consensus among voters', but also allows more analysis with recourse to simple parametric statistics.

Difficult trade-offs were faced during all phases of the survey. Selecting experts for the panel involves choosing between engaging a large more diverse set of respondents or a smaller group ensuring closer contact and higher commitment – whilst the first route was chosen for the GDPT, a substantial effort was made to keep participants engaged (e.g. with the dedicated blog). Moreover, communications with experts had to ensure adequate provision of information whilst avoiding unnecessary long messages or questionnaires that could discourage participation of experts with limited time. Finally, coding in Round 1 was challenging and time consuming. It involved two conflicting tasks: consolidation and no excessive generalization of answers - that would defeat the purpose of the Delphi.

Some limitations are inherent to the Delphi methodology. On a

procedural level, it is not possible to claim that all relevant experts were included in the panel. Databases may be incomplete, conferences, journals and other *fora* may not encompass geographically diverse members, experts may not have been recommended due to personal reasons etc. However, panel-building in the GDPT observed best practices (Delbecq et al., 1975; Marchau & van de Linde, 2016). Language may influence the outcomes of the survey: questionnaires were written in English, a possible source of difficulty for non-native speakers. This was chosen in the GDPT to guarantee that all experts were answering to the exact same questions. Particularities involving professional jargon may also have been a source of misinterpretations, both by the Delphi participants and by the coordinating team.

Content wise, the process of managing the survey involves a great deal of subjectivity. The survey coordinator has a crucial role structuring the dialogue between experts and supporting the articulation of their opinions. In the GDPT cross-checks amongst authors were used to increase the impartiality of the coordinating role when coding responses. Finally, whilst the Delphi promotes breadth, results may lack in depth. The limited time availability of experts, the need to maintain questionnaires concise, and the lack of direct interaction lead to less detailed accounts than what can be achieved with personal interviews for instance (Van Dijk, 1990). It is interesting to further develop the findings of a Delphi with additional methods that can complement the survey – workshops or case studies for instance (de Loë, 1995).

5. Conclusion and future research

Despite involving difficult trade-offs, the choice and use of the Delphi method in the article were successful, producing new and rich data. The GDPT gathered and articulated the views of diverse actors - around 60% practitioners and 40% academics, across different technical disciplines and world regions. By eliciting expert opinion to build authoritative inventories and ratings of core performance indicators and organisational features in PT, the GDPT defines *performance metrics suitable to measure strategic PT outcomes and organisational features that drive strategic outcomes*.

In relation to performance, system-wide indicators were consistently preferred by experts to assess PT. In relation to organisational features, the GDPT's results emphasise the importance of integration on policy and strategic planning levels, as well as in relation to system's tasks and components. On a more micro perspective, three **performance indicators** in particular were clearly underscored: *'User Satisfaction', 'Cost-recovery Ratio', and 'Modal Split'*, although views on the relevance of the first of them show important level of dissension. Concerning **organisational features**, policy and planning integration features such as *'Policy Integration: cross-sector links with other government areas'*, *'Integrated Planning: multimodal and multijurisdictional agency or capability'*, and *'Integrated Fare and Ticketing'* dominated the rating. Funding practices were also highlighted as relevant performance drivers, but experts' opinions in this respect were less consensual. On the other hand, 'Awarding *Mechanism Employed'* and *'Competition amongst Operating Companies'* have, according to consulted experts, a less predominant role as performance drivers.

The results produced can serve as input to future research on the dependencies and interactions amongst PT organisational and performance elements, considering the approach suggested in Section 2.3. A number of routes can be adopted to build on and complement the GDPT's findings. One alternative is to organise a workshop bringing together some of the participating experts to further discuss the results achieved, interpret the initial conclusions proposed in this article and draw new lessons. Alternatively, the same can be done via separate interviews with these experts. The direct interaction with experts, which was not possible in the GDPT, can help deepening the understanding about their views and also elucidating possible misinterpretations occurred in the online survey environment.

A third possible route is to employ the study of cases. Both across-case comparison and within-case analysis can shed light on the relationship organisation-performance and reveal how some of the features of the organisation of PT underscored in the GDPT might be connected to certain performance outcomes highlighted in the GDPT. Comparing cases might reveal contextual particularities that are also relevant and might not have been identified in the GDPT. Furthermore, the use of cases permits incorporating the longitudinal dimension of analysis, and thus to draw lessons derived from the way PT systems (and the connection organisation-performance within these systems) evolved in time. These two sets of insights (context and time) can be powerful tools for complementing the GDPT and help establishing causal links between organisation and performance in PT.

Acknowledgements

The authors are grateful for the invaluable contribution of experts who responded the survey, and for the support from colleagues A. Noroozian, C. Milchram, E. Cuppen, F. Bekius, and Q. Lone.

ANNEX I.

A. Performance Indicators Listed by Experts in the First Round of the Delphi:

SUPPLY		
Cluster	**Mentioned by**	**Experts' Description and Comments (if applicable)**
1. Vehicle km or seat km	7%	a) Total number of vehicle km or seat km (at peak hour).
2. Vehicle km ratios	9%	a) Per capita. b) Per route km/operating day. c) Per vehicle.
3. Length of network	<5%	a) Total route km per capita.
4. Operating Hours	7%	a) Service hours per capita.
COST AND REVENUE		
Cluster	**Mentioned by**	**Experts' Description and Comments (if applicable)**
5. Total Revenue and Total Cost	13%	a) Total Revenue (and total fare box revenue) and Total Cost (and total operating cost).
6. Total (and operating) costs ratios	28%	a) Per vehicle km. b) Per vehicle hour. c) Per boarding or per journey (linked trip).

(*continued on next column*)

(*continued*)

		d) Per vehicle km per hour. e) Per vehicle. f) Per capacity km. g) Per passenger km. h) Per mode. i) Maintenance cost per vehicle.
7. Labour	5%	a) Vehicle hours per employee. b) Revenue vehicle hours per operator employee hours. c) Revenue per employee. d) Vehicle km per maintenance employee. e) Revenue vehicle hours per adm. & professional employee. f) Vehicle km per employee.
8. Total (and fare box) revenue ratios	19%	a) Per vehicle km. b) Per passenger km. c) Per boarding or per journey (linked trip). d) Per line. e) Per ticket type. f) Per PT mode. g) Per area.
9. Other	5%	a) Total extra-operational revenue. b) Fuel Consumption per vehicle c) Vehicle hour per service hour.

FINANCIAL SUSTAINABILITY

Cluster	Mentioned by	Experts' Description and Comments (if applicable)
10. Cost recovery or subsidy level	37%	a) Cost-recovery. Operational revenue (fare revenue and fare substitute payments) in relation to operational costs and all costs (operating costs plus capital charge) with explicit indication of compensation payment (e.g. for concessionary travel). b) Subsidy Level. Percentage of operational costs subsidized by the government. c) Operating deficit per capita. d) Fiscal Solvency.
11. Cost per passenger ratios	24%	a) Total cost (fixed and variable) per passenger km (weighted values). b) Capital cost (equivalent average annual cost) per passenger. c) Operating cost per paid passenger km. d) Operating cost and subsidy per boarding.
12. Investment	<5%	a) Per capita spending on PT (operating and capital budgets) compared to investment in other transport modes. b) Share of urban transport investments used for PT (3 or 5 years running average).

USAGE

Cluster	Mentioned by	Experts' Description and Comments (if applicable)
13. Total Ridership	19%	a) Count of boardings and journeys (linked trips) by location and time of the day.
14. Ridership ratios	35%	a) Ridership per capita: boardings and/or journeys (linked trips) per capita. b) Total passenger km (at company or route level). c) Count of passengers by socio-economic groups.
15. Modal Split	39%	a) Ratio of PT usage (both by trips and passenger km) in relation to other modes (motorized or not) or comparison between different PT modes.
16. Occupancy Ratios (average or daily)	37%	a) Passenger per vehicle. b) Boardings per vehicle operating hour. c) Boardings per vehicle km. d) Passenger km per vehicle km. e) Passenger per vehicle km. f) Passenger km per place km (sitting and standing places). g) Passenger km per route km. h) Vehicle km (including dead running) per boarding.
17. Car ownership	<5%	a) Per capita automobile ownership and use.

SERVICE QUALITY AND USER SATISFACTION

Cluster	Mentioned by	Experts' Description and Comments (if applicable)
18. Users' access to Information	11%	a) Ease of access to information; quality and usefulness of information regarding network map, scheduled services, quality and timeliness of information on service disruptions.
19. Punctuality and Reliability	46%	a) On-time performance according to timetable. b) On-time performance based on users' perception. c) Reliability of Headways or Excess Waiting Time (train headways or average excess waiting time on selected bus routes (measure at stop level). d) Compliance to programmed offer of services. e) Breakdown Ratio: mean km before failure or breakdown rate per distance (per period). f) State of Good Repair: assets and systems are maintained to a given standard and available for service.
20. Travel Time Measures	24%	a) Ratio between travel time in PT and car. b) Average in-vehicle travel time. c) Average route time between terminal stations during peak-hour. d) Average travel times by origin and destination (in different times of the day, to achieve measure of congestion). e) Percentage of passengers with travel time up to 30 min. f) Average time waiting plus trip time plus time accessing destination. g) Total travel time divided by the total number of passengers (weighted values).

(*continued on next column*)

(*continued*)

21. Operating speed	11%	a) Average operating speed (overall and per transport mode).
22. Transfers and Connectivity	7%	a) (Average) boardings per journey for selected origin-destination pairs or measure of transfer waiting times.
23. Comfort	18%	a) Average occupation of standing passengers per square meter (during peak hour), fleet conditions (age, cleanliness etc.), and fleet characteristics (air-conditioning, Wi-Fi etc.).
24. Safety	22%	a) Ratio of staff and users injured or killed (per service-km or per trips). b) Regional per capita traffic fatality rates. c) Reach the 5% best world benchmarks.
25. Security	5%	a) Ratio of staff and users victims of crimes while in the system.
26. User Satisfaction, Acceptability and Complaints	39%	a) Index of overall user satisfaction: based on multiple criteria (punctuality, accessibility, connectivity, comfort, cleanliness, perceived safety, etc.). Measured through regular surveys with users and non-users. b) Number (and description) of complaints in relation to total number of passengers. c) Share of customized vehicles
27. Other	<5%	a) Route numbers per service hour. b) Level of Transport Infrastructure idleness.

ACCESSIBILITY IMPACTS

Cluster	Mentioned by	Experts' Description and Comments (if applicable)
28. Access to Destinations	9%	a) Number of opportunities and services that can be reached by public transport within a given time or distance.
29. Access to PT	33%	a) Average walking time or distance to access selected routes. b) Percentage of inhabitants (or users) who live within walking distance of frequent transport service. c) Number of stations or bus stops per square km. d) Distance between PT stops.
30. Fairness and Affordability	28%	a) 'Access to Destinations' measure for low income population. b) 'Access to PT' measure for the bottom 40% *'increase social inclusion and reduce inequality.'* c) Percentage of income or household budget (of low income or lowest quartile) spent on transport. d) Percentage of immobile. e) Percentage of poor served by subsidies. f) Average fare per passenger km. g) Average fare relative to petrol costs for medium-size car for short, medium and long trips (to be defined).
31. Universal Design	9%	a) Percentage of stations/stops or terminals with facilities such as on level boarding/escalators and lifts.

WIDER IMPACTS (SOCIAL, ECONOMIC AND ENVIRONMENTAL)

Cluster	Mentioned by	Experts' Description and Comments (if applicable)
32. Emissions	22%	a) Levels of air pollutants and noise nuisance per passenger km. b) Avoided CO2 emissions per passenger km in relation to CO2 emission per km by car. c) Percentage of zero or low emission vehicles in the fleet. d) Ratio of passengers transported in hybrid, electric and alternative fuel vehicles. e) Total emissions reduced in the urban transport sector and amount of reduced emissions transferred to monetary savings.
33. Energy usage	<5%	a) kWh/person-km. b) By time. c) By source.
34. Econ. Activity, Community and Regional Develop.	<5%	a) Ratio of operators' income (contract payment) reinvested in local communities and regions and not sent interstate or offshore. b) Economic activity indicators
35. Public health	<5%	a) Indicators for measuring impacts in public health and safety.
36. Accountability	<5%	a) Quality of operational data to users, communities and researchers.

OTHER

Cluster	Mentioned by	Experts' Description and Comments (if applicable)
37. Staff turnover	<5%	a) Percentage of staff turnover per passengers
38. Priority to PT	<5%	a) Share of PT network where measures to give PT priority are adopted. b) Share of infrastructure dedicated to PT in relation to that dedicated to individual modes.

B Organisational Features Listed by Experts in the First Round of the Delphi:

PT INITIATIVE AND FUNDING

Cluster	Mentioned by	Experts' Description and Comments (if applicable)
1. Initiative and Responsibility for PT	7%	a) The legal right/responsibility to initiate and regulate PT: sector is regulated and state defines entry rules or deregulated and admits market autonomous initiative. b) Level of Government with PT Responsibilities.
2. Funding Framework	30%	a) Source, availability, autonomy/control of funding for PT.
3. Commitment to PT	11%	a) Level of Political Commitment. *'In case of a higher level of political commitment the institutions will be more powerful (I hope).'* *'Transportation should be a major concern of the decision makers, and its administration trusted to competent professionals.'* b) Leadership: existence of Champion to advance PT agenda. c) Level of Investment in public transport per year. d) Adoption of public transport priority features over private modes.

(*continued on next column*)

(*continued*)

PLANNING		
Cluster	**Mentioned by**	**Experts' Description and Comments (if applicable)**
4. PT Plans	15%	a) Long-term Strategic Plan. Defining long-term quantitative and qualitative mobility goals and ensuring short-term decision-making is consistent with these goals. b) Annual Service Plan. c) The Tools for Planning and Evaluation.
5. Planning Responsibilities	41%	a) The Allocation of Planning Responsibility b) Multimodal and multijurisdictional Integration of PT Planning. Agency or Integrated Strategic Planning Capability integrating the planning of all PT modes within the metropolitan area or region and integrating all involved jurisdictions.
6. Cross-sectoral Links	22%	a) Policy Integration: Coordination between government entity responsible for PT and other government levels and policy bodies/areas (Land Use, Road, Environment, Urban Development etc.).
7. Government Entity Responsible for PT: Governance and Organisational Structure	24%	a) Procedure for Board and Management Members Selection (political appointment, election, merit-based etc.) and who it is accountable to. b) The Decision-making process. c) Concentration or Fragmentation of management structure in different layers. d) The Departmental Structure/Organisational chart indicating internal structure and responsibilities. e) Cross-organisational coordination and learning. f) The number of staff.
OPERATIONS		
Cluster	**Mentioned by**	**Experts' Description and Comments (if applicable)**
8. Operational Responsibilities	13%	a) Responsibility for Service Design, including definition of timetable, and Degree of Operational Control. b) Responsibility for Asset Management. c) The use of part-time transport operating companies for peak hours.
9. Transport Operating Companies: Ownership and Business Structure	20%	a) Nature: Private, public, mixed-capital, special-purpose company etc. b) Business structure: formal business structure *versus* informal paratransit business model. c) Annual O&M costs (size of organisation). d) Percentage of non-operational staff. e) Share of employees with pension rights etc. under no-termination contracts.
10. Operations Market Structure and Characteristics	22%	a) Competition Amongst Transport Operating Companies (existing or potential new entrants) - concentration/fragmentation of Market. b) Allocation of Ownership of long-life assets (such as garages or depots, terminals etc.). c) Size of the area over which PT is provided.
CONTRACTING PRACTICES		
Cluster	**Mentioned by**	**Experts' Description and Comments (if applicable)**
11. Tendering	39%	a) Awarding Mechanism Employed: Competitive tendering, direct award, performance-based award etc. b) Adoption of International Tendering. c) Periodic competitive tendering. d) Tendering Unit or Capability: competent team responsible for procurement of both materials from suppliers and services, applying consistent and transparent rules to all bidders. e) Transparency of the tender process and of the remuneration of transport operating companies. f) Complexity of Services to be procured. g) Involvement of Private Sector and Proportion of PT operated by private providers.
12. Contractual Regime and Elements	30%	a) Allocation of Risks and Incentive Structure. b) Use of contracts with both private and public transport operating companies. c) Length of contract with transport operating companies.
CONTROL AND ACCOUNTABILITY		
Cluster	**Mentioned by**	**Experts' Description and Comments (if applicable)**
13. Transparency of institutional setting and regulatory framework	11%	a) Clear legal and regulatory frameworks including clear mandates and performance expectations in relation to all actors.
14. Regulation	24%	a) Regulatory Agency or Capability (watchdog). b) Conflict Resolution Body: Entity responsible for conflict resolution between government entity responsible for PT and transport operating companies. c) Use of Price Regulation. d) Pro-active law enforcement - criminal and traffic related incidents.
15. Social Control	>5%	a) Consulting Forum for Control of Management. b) Visibility of elected officials.
16. User Orientation	17%	a) Community and User Outreach: Established mechanisms for consultation and communication with community and users. b) 'Urban Mobility Observatory'.
17. Budget constraints	>5%	a) Adoption of Budget Constraints
OTHER INTEGRATION AND FRAGMENTATION FEATURES		
Cluster	**Mentioned by**	**Experts' Description and Comments (if applicable)**
18. System Integration	35%	a) Integrated fare and ticketing (clearinghouse functions). b) Degree of Vertical Integration (Infrastructure and Operations).

(*continued on next column*)

(*continued*)

		c) Integrated information system. d) Integration of Feeder services and connection times. e) Ability to interact with industry (entities such as taxi companies and bike sharing).
19. Other Integrating Bodies	<5%	a) Use of Infrastructure Coordinator. b) Use of Financial Authority.
20. Separation of Responsibilities	20%	a) Separation of Planning and Regulating responsibilities. b) Separation of Planning and Operating responsibilities. c) Separation of Planning and Funding Responsibilities. d) Separation of Funding and Operating responsibilities.

PEOPLE AND RELATIONSHIPS

Cluster	Mentioned by	Experts' Description and Comments (if applicable)
21. Staff Elements	28%	a) Skill set and technical expertise of staff. b) Professional development: programs for recruiting, training, and retaining staff. c) Degree of gender diversity in staff of transport operating companies. d) Workforce relations.
22. Trust, Partnership and Communication	11%	a) Trust/Partnership/Communication between government entity responsible for PT and transport operating companies. b) Degree to which information is shared amongst transport operating companies.

OTHER

Cluster	Mentioned by	Experts' Description and Comments (if applicable)
23. Fare Setting, Review, and Collection	11%	a) Allocation of fare tasks; fare practices.
24. Level of innovativeness	7%	a) Ability of government and operating companies to incorporate innovative practices and new technologies, including the use of information technology systems.

References

Albalate, D., Bel, G., & Calzada, J. (2012). Governance and regulation of urban bus transportation: Using partial privatization to achieve the better of two worlds. *Regulation and Governance, 6*(1), 83–100.

Banister, D. (2008). The sustainable mobility paradigm. *Transport Policy, 15*, 73–80.

Barter, P. (2008). Public planning with business delivery of excellent urban public transport. *Policy and Society, 27*, 103–114.

Beck, A. (2011). Experiences with competitive tendering of bus services in Germany. *Transport Reviews, 31*(3), 313–339.

Bowen, G. A. (2006). Grounded theory and sensitizing concepts. *International Journal of Qualitative Methods, 5*(3), 12–23.

Brown, J. R., & Thompson, G. L. (2008). Service orientation, bus-rail service integration, and transit Performance: Examination of 45 U.S. Metropolitan areas. *Transportation Research Record: Journal of the Transportation Research Board, 2042*, 82–89.

de Bruijn, H., & Dicke, W. (2006). Strategies for safeguarding public values in liberalized utility sectors. *Public Administration, 84*(3), 717–735.

Buehler, R., & Pucher, J. (2011). Making public transport financially sustainable. *Transport Policy, 18*(1), 126–138.

Chadwick, E. (1859). Results of different principles of legislation and administration in europe; of competition for the field, as compared with competition within the field, of service. *Journal of the Statistical Society of London, 22*(3), 381–420.

Chisholm, D. (1992). *Coordination without Hierarchy: Informal structures in multiorganizational systems*. University of California Press.

Cowie, J. (2014). Performance, profit and consumer sovereignty in the English deregulated bus market. *Research in Transportation Economics, 48*, 255–262.

Cuppen, E. (2010). *Putting perspectives into Participation: Constructive conflict methodology for problem structuring in stakeholder dialogues*. Uitgeverij BOXPress.

Dajani, J. S., & Gilbert, G. (1978). Measuring the performance of transit systems. *Transportation Planning and Technology, 4*(2), 97–103.

Dalkey, N., & Helmer, O. (1963). An experimental application of the Delphi method to the use of experts. *Management Science, 9*(3), 458–467.

Delbecq, A. L., Van de Ven, A. H., & Gustafson, D. H. (1975). *Group techniques for program planning: A guide to nominal group and delphi processes*. Scott, Foresman and Company.

Docherty, I., Shaw, J., & Gather, M. (2004). State intervention in contemporary transport. *Journal of Transport Geography, 12*, 257–264.

Faivre d'Arcier, B. (2014). Measuring the performance of urban public transport in relation to public policy objectives. *Research in Transportation Economics, 48*, 67–76.

Farrington, J. H. (2007). The new narrative of accessibility: Its potential contribution to discourses in (transport) geography. *Journal of Transport Geography, 15*, 319–330.

Fielding, G. J. (1992). Transit performance evaluation in the U.S.A. *Transportation Research Part a, 26A*(6), 483–491.

Fielding, G. J., Babitsky, T. T., & Brenner, M. E. (1985). Performance evaluation for bus transit. *Transportation Research Part a: General, 19A*(1), 73–82.

Finn, B., & Mulley, C. (2011). Urban Bus Services in Developing Countries and Countries in Transition: A Framework for Regulatory and Institutional Developments. *Journal of Public Transportation, 14*(4), 89–107.

Geerlings, H., Klementschitz, R., & Mulley, C. (2006). Development of a methodology for benchmarking public transportation organisations: A practical tool based on an industry sound methodology. *Journal of Cleaner Production, 14*(2), 113–123.

Gómez-Lobo, A. (2007). Why competition does not work in urban bus markets: Some new wheels for some old ideas. *Journal of Transport Economics and Policy, 41*(2), 283–308.

Hale, C. A. (2011). New approaches to strategic urban transport assessment. *Australian Planner, 48*(3), 173–182.

Hensher, D. A., & Wallis, I. (2005). Competitive Tendering as a contracting mechanism for subsidising transport: The Bus Experience. *Journal of Transport Economics and Policy, 39*(Part 3), 295–321.

Jain, P., Cullinane, S., & Cullinane, K. (2008). The impact of governance development models on urban rail efficiency. *Transportation Research Part a: Policy and Practice, 42*, 1238–1250.

Jørgensen, T. B., & Bozeman, B. (2007). Public Values An Inventory. *Administration & Society, 39*(3), 354–381.

Kavanagh, P. (2016). A case for negotiated performance-based contracting rather than competitive tendering in government public transport (bus) service procurement. *Research in Transportation Economics, 59*, 313–322.

Linstone, H. A., & Turoff, M. (1975). The Delphi Method - Techniques and applications. In H. A. Linstone, & M. Turoff (Eds.), *The Delphi method - techniques and applications*. Addison-Wesley Educational Publishers Inc.

de Loë, R. C. (1995). Exploring complex policy questions using the policy Delphi. A multi-round, interactive survey method. *Applied Geography, 15*(1), 53–68.

Marchau, V. A. W. J., & van de Linde, E. (2016). The Delphi Method. In P. van der Duin (Ed.), *Foresight in organizations: Methods and tools* (pp. 59–79). Routledge.

Marshall, S. (2001). The challenge of sustainable transport. In A. Layard, S. Davoudi, & S. Batty (Eds.), *Planning for a sustainable future* (pp. 131–147). Taylor & Francis.

Ostrom, E. (1990). *Governing the Commons: The evolution of institutions for collective action*. New York: Cambridge University Press.

Ostrom, E. (2010). Beyond Markets and States : Polycentric Governance of Complex Economic Systems. *The American Economic Review, 100*(3), 641–672.

Paredes-Molina, R., & Baytelman, Y. (1996). Urban Transport Deregulation: The Chilean Experience. *Estudios de Economia, 23*, 195–210.

Ragin, C. (1987). *The comparative Method: Moving beyond qualitative and quantitative strategies*. University of California Press.

Roy, W., & Yvrande-Billon, A. (2007). Ownership, contractual practices and technical efficiency: The case of urban public transport in France. *Journal of Transport Economics and Policy, 41*(2), 257–282.

Schmidt, R. C. (1997). Managing Delphi Surveys Using Nonparametric Statistical Techniques. *Decision Sciences, 28*(3), 763–774.

Stanley, J., & Hensher, D. A. (2008). Delivering trusting partnerships for route bus services: A Melbourne case study. *Transportation Research Part a: Policy and Practice, 42*(10), 1295–1301.

Thacher, D., & Rein, M. (2004). Managing Value Conflict in Public Policy. *Governance: An International Journal of Policy, Administration and Institutions, 17*(4), 457–486.

Van Dijk, J. A. (1990). Delphi questionnaires versus individual and group interviews. *Technological Forecasting and Social Change, 37*(3), 293–304.

Veeneman, W. (2002). *Mind the gap: Bridging theories and practice for the organisation of metropolitan public transport*. Delft University Press.

Veeneman, W., Augustin, K., Enoch, M., Faivre d'Arcier, B., Malpezzi, S., & Wijmenga, N. (2015). Austerity in public transport in Europe: The influence of governance. *Research in Transportation Economics, 51*, 31–39.

van de Velde, D. M. (1999). Organisational forms and entrepreneurship in public transport. Part 1: Classifying organisational forms. *Transport Policy, 6*(3), 147–157.

van de Velde, D. M., Nash, C., Smith, A. S., Mizutani, F., Uranishi, S., Lijesen, M., et al. (2012). *Eves-Rail - economic effects of Vertical Separation in the railway sector; Full technical report for CER*. Amsterdam: Community of European Railways and Infrastructure Companies.

van de Velde, D. M., & Wallis, I. (2013). "Regulated deregulation" of local bus services-An appraisal of international developments. *Research in Transportation Economics, 39*(1), 21–33.

Walker, J. (2008). Purpose-driven public transport: creating a clear conversation about public transport goals. *Journal of Transport Geography, 16*, 436–442.

Zhang, C., Juan, Z., & Xiao, G. (2015). Do contractual practices affect technical efficiency? Evidence from public transport operators in China. *Transportation Research Part E: Logistics and Transportation Review, 80*, 39–55.