

Experts' Inventory and Rating of Core Performance Indicators and Organisational Features in Public Transport: A Global Delphi Survey

Fabio Hirschhorn^{a*}, Wijnand Veeneman^a, Didier van de Velde^a

^a*Faculty of Technology, Policy and Management, Delft University of Technology*

*Corresponding author: Jaffalaan 5, 2628BX, Delft, The Netherlands, +31 15 2784885, f.hirschhorn@tudelft.nl

Keywords: Public Transport - Performance Indicators - Organisational Form - Delphi Survey

Classification codes: H440, L510, L980, R000

ABSTRACT

Discussions involving the connection between organisational form and performance in public transport systems take place for over a century and remain at the top of the agenda in the sector. These analyses indicate that some organisational elements might be important for success in public transport. However, studies usually focus on the impacts of a single policy initiative and results are seldom clear. The analysis of the relationship organisation-performance can benefit from a more comprehensive approach that does not treat variables only in an isolated and incremental fashion, but instead considers performance outcomes as the result of the combined effect of different organisational variables. A first step to allow this broader approach is to get the variables right: *What organisational features drive strategic performance outcomes, and what performance metrics are more suitable to measure these impacts?* A Delphi survey is developed and, based on experts' judgement, builds authoritative lists of core performance indicators and organisational features. User Satisfaction and Policy Integration between Public Transport and other government areas emerge, respectively, as highest rated variables.

Keywords: Public Transport - Performance Indicators - Organisational Form - Delphi Survey

1. INTRODUCTION

How is the performance of metropolitan public transport ("PT") influenced by its organisational setting? The connection between performance and organisation in PT has been under analysis for over a century (the discussion in Chadwick (Chadwick, 1859) illustrates this) and remains at the top of the academic agenda (Faivre d'Arcier, 2014; Hensher & Wallis, 2005). These analyses indicate that some organisational features (market deregulation, use of competitive tendering etc.) might be important for achieving goals attached to PT (efficiency, accessibility etc.).

However, unveiling such relationship is not simple and studies in this field have mixed or inconclusive results (Roy & Yvrande-Billon, 2007; van de Velde & Wallis, 2013). Difficulties are multiple. The identification of suitable performance and organisational variables to be examined is the first of them.

Concerning **performance**, two levels of analysis exist: the first refers to broad PT aims (examples above). Priorities in relation to broad aims change over time and according to the varied (sometimes conflicting) expectations actors have in relation to PT. The second level is related to the translation of broad aims into quantitative metrics - performance indicators

(vehicle hour per employee, population living within 500m to frequent PT service etc.). The choice of the best indicator or set of indicators to measure broad aims varies substantially and consensus is a challenge (Fielding, 1992). With regards to **organisational form**, difficulties emerge as these features are ultimately shaped by local characteristics. Any analysis of their influence over performance requires in depth research of complex particularities, limiting the generalization of conclusions.

Consequently, the study of the relationship organisation-performance is often driven by the assessment of impacts of one specific policy creating or reforming a feature of PT organisation. Impacts are assessed according to particular goals that are/were to be implemented and these change from place to place. As a result, only a direct and simple relationship between isolated organisational features and performance indicators is examined over time, pre and post implementation of changes. This approach produces excellent research, yet it only allows an incremental analysis of policies and does not acknowledge PT as a complex system in which (i) multiple stakeholders interact, (ii) diverse public values coexist and (iii) variables affect one another. Performance-related outcomes may result from the combination of different configurations of organisational variables.

A first step to approach the relationship organisation-performance more comprehensively is to identify research variables that enable an analysis not guided by the assessment of a single policy. For the reasons exposed above, academic literature alone is not sufficient to offer this. This article resorts to experts and establishes authoritative lists of core (i) performance indicators and (ii) organisational features driving performance in PT. This is done with a Global Delphi in Public Transport (“GDPT”), following an iterative three-phase process in which questionnaires are interspersed with information and opinion feedback.

The article first reviews academic literature on performance and organisational form (Section 2). In continuation, the Delphi methodology and the GDPT are described (Section 3). Outcomes are discussed (Section 4), followed by conclusions (Section 5).

2. BACKGROUND LITERATURE

2.1. Performance

PT performance involves two levels of analysis. The first relates to broad dimensions or aims that are attached to PT, such as efficiency, accessibility etc. The second relates to the quantitative measures through which broad performance aims can be translated – performance indicators. Efficiency, for instance, could possibly be measured by annual operating deficit, employee per vehicle hour etc. Performance aims can be associated to an *ex-ante* focus on policy design and interventions, whereas performance indicators have an *ex-post* focus on outcomes.

Broad aims change in time and according to actors’ interests. Literature 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; Geurs & van Wee, 2004; Preston & Rajé, 2007). More recently, an interest in user satisfaction as a goal is also emphasised (Fiorio, Florio, & Perucca, 2013; Mouwen & Rietveld, 2013).

Aims may also compete. Walker (2008) distinguishes patronage goals - those achieved by the extent people use PT - and coverage goals - concerned with availability of PT. He 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).

Concerning performance indicators, the choice of indicator or set of indicators to measure performance varies and consensus is a challenge (Fielding, 1992). Firstly, the range of options is vast. Geerlings et al. (2006) report that the literature review undertaken as part of the EU project EQUIP identified over 400 indicators in PT. Additionally, 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) proposes 5 efficiency metrics to be adopted by USA PT agencies (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 et al. (2008) employ Data Envelopment Analysis to compare technical efficiency in 15 PT systems measuring vehicle kilometres and also passenger trips (supply and demand-oriented metrics).

2.2. Organisational Form

Different authors characterize the organisational structure of PT (Barter, 2008; Grant, Pollan, & Blake, 2011; van de Velde, 1999). Based on these models, it is possible to identify some of PT organisational features: existence of a legal monopoly for the government to set-up and regulate public transport services (as opposed to deregulated regimes with autonomous market entry), division of regulatory powers between government tiers and agencies, funding practices, ownership structure, competition 'for' the market, different awarding mechanisms etc.

The contrast between markets with open entry to autonomous players and markets where the public sector holds a legal monopoly to initiate and organize PT services constitutes an important research track in PT. Chadwick (1859) examines competition for the field and competition within the field, a theme that again spurred research after more recent deregulation experiences in the UK and elsewhere (Cowie, 2014; Paredes-Molina & Baytelman, 1996).

Impacts of regulated competition are researched in various PT markets (Gómez-Lobo, 2007; Zhang, Juan, & Xiao, 2015). Different awarding mechanisms and contractual regimes between government authorities and operating companies also constitute research avenues (Kavanagh, 2016; Stanley & Hensher, 2008).

Authors examine ownership issues and effects of vertical integration or separation (Scheffler, Hartwig, & Malina, 2013). At the level of transport operating companies, the performance of private and corporatized transport undertakings is compared (Albaladejo, Bel, & Calzada, 2012). In the USA private operating companies are less present; public entities that accumulate the roles of planning authority and operating company are more common (Grant et al., 2011).

The tier of government responsible for PT, key stakeholders and their interaction are also studied. Henkin et al. examine PT agencies in the USA and Canada to identify transformative institutional changes (Henkin et al., 2012). Buehler & Pucher (2011) highlight regional coordinating authorities in Germany, whilst Finn & Mulley (2011) discuss the role of the market regulator. Finally, implications of funding policies (Veeneman et al., 2015) and different service characteristics (Brown & Thompson, 2008) are also analysed in PT literature.

2.3. Performance and Organisation

Analyses described in Sections 2.1 and 2.2 offer important insights. The point is not to judge whether there are higher aims in PT, single organisational features to achieve them, or indisputable metrics to assess outcomes. The upshot is that when driven by the assessment of a single policy, studies examine variables in an isolated manner. Thus, academic literature misses some of the complexities of the mechanisms connecting organisational form and performance.

PT is a multifarious system in which numerous actors with diverse and conflicting values coexist. Competing goals and changing priorities have to be realised, and several strategies to operationalize and safeguard them take place (Thacher & Rein, 2004). To acknowledge this complexity, it is necessary to search for alternative ways to analyse PT. This does not mean leaving aside research being done, but rather adding new analytical approaches to current practice.

This article proposes that the connection between organisational form and performance can be better understood as a configuration of interacting elements instead of the summation of isolated policy interventions. The effect of variables is not independent from one another and the relationship organisation-performance is complex in at least three ways: first, performance outcomes can result from the combination - at the same time or in sequence - of different organisational variables. Second, different combinations of conditions may produce the same outcome. Third, a given condition may have different impacts on outcomes depending on context (Rihoux & Lobe, 2009).

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). He suggests that one necessary step is to find metrics that can help measuring strategically important transport initiatives or projects, metrics 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.” (2011, p. 175).

This article's purpose is similar. It recognizes the complexity of PT and, to examine the relationship organisation-performance with more comprehensive lens, it aims at defining *what organisational features drive strategic performance outcomes, and what performance metrics are more suitable to measure these impacts*. A Delphi survey 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 opinion feedback to build authoritative forecasts in relation to the occurrence of events or trends (Dalkey & Helmer, 1963). After the first questionnaire, and preceding new rounds, participants have access to and can reflect on other experts' opinions through anonymous feedback organized by the survey's coordinator.

Variants of Delphi emerged promoting different outputs. The Policy Delphi, for instance, '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 can be used to establish the relative importance of issues building authoritative rankings (Schmidt, 1997).

The Delphi can help answering complex issues that cannot rely on the knowledge of a single expert or group of stakeholders. The anonymity of the process supports free expression of opinions. Furthermore, the method circumvents the practical difficulty of bringing all experts together. Finally, the method permits engaging different types of stakeholders, moving beyond academic boundaries.

3.2. The Global Delphi in Public Transport

3.2.1. Structure and Experts

The GDPT had three stages: brainstorming (respondents could freely propose relevant elements in connection to the issues at stake), narrowing-down (respondents shortlisted most relevant elements from previous stage) and rating (respondents rated shortlisted elements). Each stage used one questionnaire.

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

Concerning the relevant expertise, the GDPT attempted to cover knowledge on (i) PT performance monitoring and evaluation, and (ii) PT governance - design and functioning of organisational forms adopted in diverse systems. In addition, three other aspects were considered: first, including individuals of prominence in the field. Second, blending stakeholders with different roles in PT, ensuring access to multiple views. Finally, including experts with knowledge on varied PT organisational settings.

To identify experts covering these characteristics, the following criteria were used:

- Prominence: affiliation to eminent organisations and involvement in major international *fora* (e.g. Transportation Research Board), major universities, government entities responsible for PT, PT operating companies, participation in editorial boards of prominent journals (Transport Policy, Transportation etc.).
- Variety of roles: representatives from (i) academia, (ii) government entities responsible for PT, (iii) PT operating companies (public and private); (iii) users' associations, (v) multilateral institutions, (vi) consultancy. Regarding academics, two more aspects were considered: publications (retrieved on Google Scholar), and variety of academic discipline: transport geography; transport economics; transport engineering; public administration and policy; and urban planning (identified by Veeneman (Veeneman, 2002)).
- Variety of organisational settings: experts based in, and with expertise on, varied geographical regions.

A first search resulted in 170 names that also included contacts from the authors. Some initially identified experts were asked to recommend experts – suggested names that had not been identified previously were included in the panel. Over 200 experts were identified. They were then assessed more closely to confirm their involvement with the survey's topics. Eventually 169 experts were invited to the GDPT. Considering invitations that bounced back and accesses to the questionnaire's online platform, it is possible to confirm that 96 experts received this invitation.

3.2.2. Brainstorming

The first questionnaire of the Delphi had two types of questions. First, in open-ended questions, experts were asked to assume they would be hiring a consultant to analyse a PT system and, thus, had to list and describe at least 5 performance indicators they would want examined. The selection had to prioritize indicators better able to provide relevant information on critical aspects of PT (without pre-determining what critical meant). Experts were also asked to associate each indicator to a broad aim. A similar question asked experts to list and describe at least 5 organisational features considered important drivers of PT performance (without defining a specific type of performance) to be examined by the consultant.

In a second set of questions, experts were presented with and asked to rate the relevance of some 'pairs' of performance aims and indicators normally present in PT literature (e.g. *Environmental Sustainability; per capita emissions of NOx*). The same rating was done with a list of organisational features (e.g. *Allocation of Ownership of Long-Life Assets*). These rating questions were '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. A total of 54 experts completed questionnaire one.

Answers to open-ended questions were qualitatively analysed. Major themes were identified, redundancies eliminated, and a long inventory of performance indicators and organisational features was produced. The 'rating safety-net' was not needed.

In relation to **performance**, existing evaluation frameworks (1978; Fielding, 1987) were taken as source of sensitizing concepts - a general guideline for the interpretation and

organisation of input received from experts (Bowen, 2006). These evaluation schemes define three classes of performance measures (i) *cost-efficiency* (technical relationships between service inputs 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, *impact measures* 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 though, and the inventory assumed its own form, shaped by actual responses received.

The attempt to have broad aims associated to indicators highlighted how controversial the issue is. Same indicators were assigned to different aims by respondents. Furthermore, this conceptual discussion was mostly restricted to academics - practitioners in the panel hardly used definitions as efficiency, effectiveness etc. when responding to the part of the question asking for this. The decision was not to persist with the discussion of broad aims in following Rounds, lest participants could lose interest in the survey.

The following broad performance themes were identified: '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.

Table 1: Sample from Performance Indicator Inventory

ACCESSIBILITY IMPACTS		
Cluster	Mentioned by	Indicators
Access to Destinations	9%	a) Number of opportunities and services that can be reached by public transport within a given time or distance.
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.
Fairness and Affordability	28%	a) 'Access to Destinations' measure <u>for low income population</u> . b) 'Access to PT' measure <u>for the bottom 40%</u> '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).
Universal Design	9%	a) Percentage of stations/stops or terminals with facilities such as on level boarding / escalators and lifts.

Note: the full inventories can be made available upon request to authors.

Concerning **Organisational Features**, van de Velde (1999) was used as a source of sensitizing 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. This framework was not employed prescriptively and responses provided by experts shaped the final inventory.

The convoluted use of terminology in PT became evident. 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 and the use of the same denomination causes misunderstandings. Additionally, in this Round 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.

The coding of responses identified the following major themes: '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 that contained a total of 70 organisational features.

The feedback material sent to experts after Round 1 included the full inventory of variables, indication of the percentage of respondents that mentioned each cluster (Table 1), and all comments from respondents.

3.2.3. Narrowing-down

The shortlisting exercise required experts to select 7 performance indicators and 7 organisational features amongst all those previously listed - this was the average number of performance indicators listed by each respondent in Round 1. This selection was again supposed to prioritize indicators better able to provide insights on most critical aspects of PT and organisational features with greater impact on performance. A total of 48 experts completed Round 2. A few respondents selected more than 7 items. They had their votes considered based on weighted values.

Although questions asked for the selection of specific individual variables, the analysis of responses also took into account what would have been the sum of votes given to each of the clusters defined in Round 1. This revealed that a shortlist of 7 items would not adequately reflect experts' priorities. In particular, the third most voted cluster would have been *Total (and operating) costs ratios*, but these votes were dispersed amongst 9 alternative indicators. Whilst clearly relevant to experts, none of them would have been included in the third stage of the GDPT with a shortlist of 7 items. Two shortlists, each with 10 variables, were taken to Round 3, based on simple majority of votes (Schmidt, 1997) (Table 2).

Some experts mentioned the lack of a specific policy objective to guide the choice of 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*'. Attaching experts' choices in the GDPT to the assessment of specific policies and goals was avoided to attempt following a more comprehensive approach to PT (Section 2.3).

Feedback material sent to respondents described the shortlists, the ratio of votes each variable received, and included comments from experts.

Table 2: Summary results Round 2

Performance Indicators	Organisational Features	Voted by
Cost-recovery Ratio	N/A	45%
Modal Split	N/A	41%
User Satisfaction (overall index)	N/A	37%
On-time Performance According to Timetable	N/A	31%
Ratio between Travel Time in PT and Car	N/A	30%
% of Inhabitants (or users) living within walking distance to Frequent PT Service	N/A	30%
Ridership per Capita	N/A	27%
Cost per Passenger Km	N/A	21%
Total Revenue and Total Cost	N/A	17%
Comfort	N/A	16%
N/A	Integrated Fare and Ticketing (clearinghouse functions)	59%
N/A	Funding: source, availability, autonomy/control	57%
N/A	Integrated Planning: multimodal and multijurisdictional agency or capability	43%
N/A	Policy Integration (cross-sector links with other govt. areas)	39%
N/A	Clear Legal and Regulatory Frameworks	34%
N/A	Long-term Strategic PT Plan	31%
N/A	Contracts: risk allocation and incentive structure	28%
N/A	Skill set and technical expertise of staff	27%
N/A	Awarding Mechanism Employed	23%
N/A	Competition among Operating Companies (re. market concentration/fragmentation)	21%

Note: 'N/A'=non-applicable.

3.2.4. Rating

In the final questionnaire experts were asked to allocate a total of 100 points amongst (all or part of) the variables in each shortlist. Points were to be freely distributed to reflect relevance of variables if they were to be used for a hypothetical comparative study of PT in 15 metropolitan areas. The stated objective of this study was to unveil how different ways of organizing PT influence different types of performance. Wording once more referred to indicators' ability to provide insights on strategic aspects of PT and organisational features' ability to drive performance. Variables were randomized in the questionnaires to avoid inducing bias. In total, 45 experts completed Round 3 (Table 3).

4. DISCUSSION

4.1. Experts and Method

The GDPT grouped a diverse panel (Figure 1) and promoted a qualitative exercise - by design not statistically significant - with results that could hardly be achieved differently. The periodic feedback helped in keeping experts engaged and the high response rates testify to this: in addition to partial results, a blog was used to post survey's details and updates. The survey combined elements of a Policy Delphi (multiple and sometimes conflicting views on issues) and a ranking-type Delphi (building authoritative lists). By introducing the point allocation methodology instead of a simple ranking question in the last Round, GDPT's analysis are not restricted to measuring consensus and can be expanded with recourse to simple parametric statistics.

Difficult trade-offs were faced. Selecting experts for the panel required choosing between engaging a large more diverse set of respondents or a smaller group ensuring closer contact and higher commitment. Moreover, communications 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 involved two

conflicting tasks: consolidation and no excessive generalization of answers - that would defeat the purpose of the Delphi.

The GDPT also has limitations. First, it is not possible to claim that all relevant experts were included in the panel, or that the final list has no bias. Databases used 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). Furthermore, language may influence the outcomes of the survey: questionnaires were written in English, a possible source of difficulty for non-native speakers. This was preferred to guarantee that all experts were answering to the exact same questions. Finally, whilst the Delphi promotes breadth, results may lack in depth. The need to maintain questionnaires concise and the lack of direct interaction may create shortcomings (Van Dijk, 1990).

4.2. Results

This section takes stock on outcomes of Rounds 2 and 3, and also changes between Rounds. Whilst already interesting, this assessment is not exhaustive and only highlights a first catalogue of remarks that warrant further investigation.

Concerning **performance indicators**, Round 2 allows the following observations:

- Some shortlisted indicators interact or overlap, e.g. '*User Satisfaction*' and '*Comfort*' (the latter is a component of the second), same with '*Total Revenue and Total Cost*' and '*Cost per Passenger Km*'. Rather than a problem with the survey, these 'redundancies' demonstrate some degree of 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. They also reflect controversies on the choice of indicators to translate aims (Section 2.1). Therefore, assessing relative importance of variables in Round 3 adds an important analytical layer.
- The shortlist shows prevalence of indicators associated to patronage goals (as defined by Walker (Walker, 2008)). Only '*% of inhabitants (or users) living within walking distance to frequent PT service*' is related to coverage goals.
- 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.
- Amongst absences from the shortlist, '*Vehicle Km per Capita*' and '*Subsidy Level*' had voting rates close to the 'top 10' items. A limit for the shortlist had to be set though. Furthermore, no indicator of environmental impact or affordability was shortlisted. The 'best ranked' amongst them was '*Percentage of income or household budget spent on transport*', the 26th in the overall list. This is surprising considering the increasing attention these aspects receive in literature and policy *fora*.

Results from Round 3 add new angles:

- When asked to consider the relative importance of shortlisted indicators, some participants realised interactions and overlaps produced in Round 2. E.g. '*Some of these 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.

- The three most voted indicators remained the same between Rounds. The upshot in Round 3 is the gap in points separating them from the rest of the list. This underscores what may be seen as three key PT performance indicators.
- ‘*User Satisfaction*’ tops the list, maybe denoting the rise in importance given to the users’ perspective (Section 2.1) and possibly how useful this measure can be for service evaluation and design. Surprisingly, PT research rarely engages directly with users being often restricted to analysing results of customer surveys. Experts’ views on ‘*User Satisfaction*’ are highly dispersed (standard deviation equals 11,0).
- Opinions around ‘*Comfort*’ are less dispersed. They present the lowest standard deviation (5,9) and a high percentage of ‘zeros’ (39%). This may be caused by the overlap with *User Satisfaction*. The same may explain the variation in ranking between Rounds of ‘*On-time Performance according to Timetable*’.
- The very high percentage of ‘zeros’ to ‘Total Revenues and Total Costs’ (43%) corroborates opinions expressed by some experts advocating that cost measures are more informative per unit of production. Its presence in the shortlist can be seen as contradictory, and might be a reflection of terminology misinterpretations during the Delphi.

Regarding **Organisational Features**, Round 2 highlights:

- Interactions and overlaps amongst features are also present. They reflect experts’ emphasis on PT planning tasks and also financial practices in the sector.
- Experts also stressed system integrative features (‘*Integrated Fare and Ticketing*’, ‘*Integrated Planning*’, ‘*Policy Integration (cross-sector links with other govt. areas)*’). Although integration through hierarchic governance is not a condition for coordination, either in PT (Chisholm, 1992) or more generally (Ostrom, 1990), the rationale associating both seems to have guided some choices. E.g. ‘*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.*’
- Most shortlisted features involve tasks in the Strategic and Tactical levels of PT (as defined by van de Velde (van de Velde, 1999)).
- ‘*Integrated Fare and Ticketing*’ and ‘*Funding: source, availability, autonomy/control*’ had, by far, more votes, showing some consensus amongst experts about the relevance of these features as performance drivers. More interesting, perhaps, is the relative consensus on ‘*Awarding Mechanism Employed*’ and ‘*Competition amongst Operating Companies*’ as less important drivers of performance. It is true that they are shortlisted, however one could expect them to be ‘leading the polls’ given their prominence in academic literature. Have experts concluded these features are less impacting than previously imagined?
- ‘*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 at the centre of recent policy and academic debates. Nonetheless, they are not shortlisted.

Other interesting aspects involving organisational features emerge in Round 3:

- Consistently with Round 2, PT planning tasks and integrative features are emphasised by experts. Interactions amongst features may influence this, as seen from some comments. E.g. ‘*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.’

- ‘Policy Integration (cross-sector links with other govt. areas)’ topped the ranking, whilst ‘Integrated Fare and Ticketing’ moved to the fifth position. Many hypotheses can be formulated to explain this. One possibility is the framing of the question proposing the use of features for a comparative study of metropolitan areas – maybe experts consider fare and ticketing integration less informative if used for this purpose.
- A gap in points between features is clear, this time separating the list in two blocks of five variables. This underscores that to experts two organisational dimensions are key - planning and funding - and an organisational characteristic is key as well – integration.
- Another visible gap separates the last two ranked features from the rest of the list. These are ‘Awarding Mechanism Employed’ and ‘Competition amongst Operating Companies’. Votes for these two features less dispersed (low standard deviation) and, most strikingly, they received ‘zeros’ from nearly 50% of respondents.

Table 3: Summary results Round 3

Performance Indicators	Avg. Points	Std. Dev.	High	Mode	Zeros	Rank	Rank 2
User Satisfaction (overall index)	16,3	11,0	50	15	9%	1	3
Cost-Recovery Ratio	14,9	9,0	30	20	14%	2	1
Modal Split	13,5	9,8	40	20	20%	3	2
% of Inhabitants (or users) living within walking distance to Frequent PT Service	9,7	7,7	30	10	23%	4	6
Ridership per Capita	9,4	8,7	30	5	25%	5	7
Ratio between Travel time in PT and Car	8,3	6,3	25	10	23%	6	5
On-time Performance according to Timetable	7,7	6,9	20	0	32%	7	4
Total revenue and total cost	7,4	8,5	30	0	43%	8	9
Cost per Passenger Km	7,0	6,7	26	0	34%	9	8
Comfort	5,9	5,9	20	0	39%	10	10

Organisational Features	Avg. Points	Std. Dev.	High	Mode	Zeros	Rank	Rank 2
Policy Integration (cross-sector links with other govt. areas)	14,6	7,9	30	10	9%	1	4
Funding: source, availability, autonomy/control	13,5	10,1	50	15	18%	2	2
Integrated Planning: multimodal and multijurisdictional agency or capability	12,8	8,4	40	10	14%	3	6
Long-term Strategic PT plan	12,7	7,9	30	10	11%	4	3
Integrated Fare and Ticketing (clearinghouse functions)	12,5	8,8	40	10	23%	5	1
Clear Legal and Regulatory Frameworks	9,1	6,9	30	10	23%	6	5
Contracts: risk allocation and incentive structure	8,5	7,6	30	10	27%	7	7
Skill set and technical expertise of staff	7,7	7,0	25	0	32%	8	8
Awarding Mechanism Employed	4,6	5,2	20	0	45%	9	9
Competition among Operating Companies (re. market concentration/fragmentation)	4,0	4,5	15	0	48%	10	10

Note: ‘Avg. Points’ is the average number of points allocated to the variable by each expert; ‘High’ indicates the highest punctuation attributed to the variable by one expert; ‘Zeros’ shows the percentage of experts attributing zero point to a variable; ‘Rank’ ranks variables based on the amount of points they receive; ‘Rank 2’ shows variables vote ranking in Round 2.

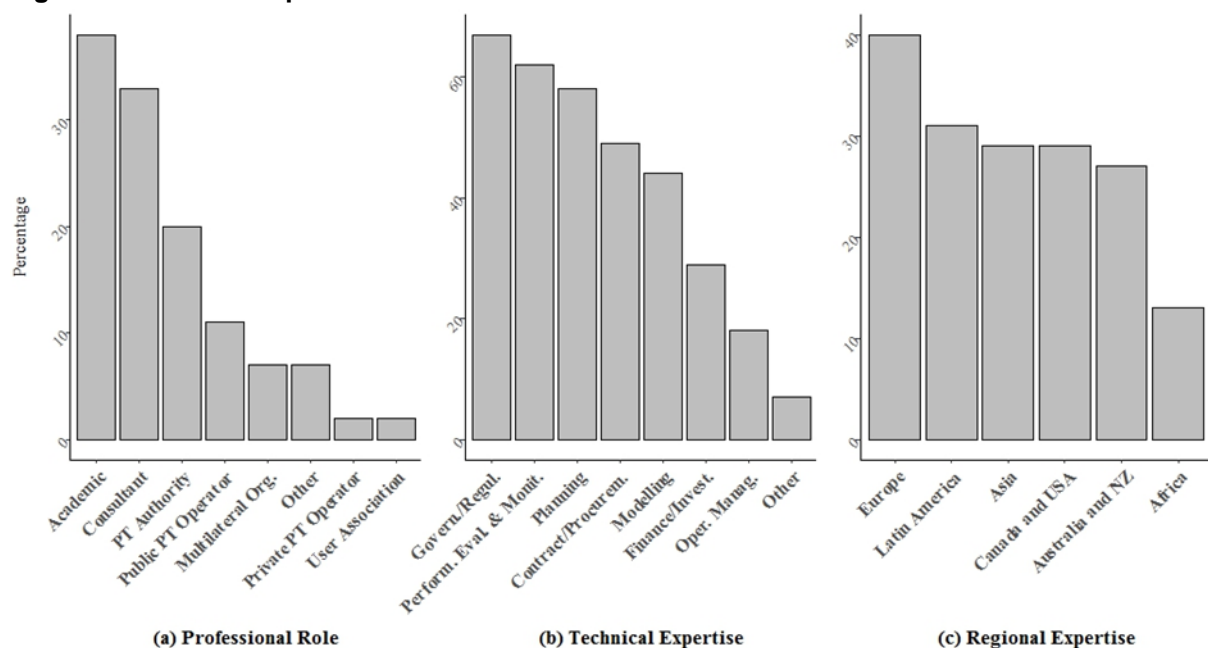
5. CONCLUSION

Despite involving difficult trade-offs and promoting less depth than breadth, the choice and use of the Delphi method for this research was successful. The methodology is flexible and allows the production of rich data that would hardly be achieved differently. Furthermore, the GDPT was effective in gathering the views of multiple actors, beyond academia, and in

building authoritative inventories and ratings of core performance indicators and organisational features in PT.

Multipart system-wide performance indicators connected to patronage goals were prioritized as better able to provide insights on strategic aspects of PT. Three key indicators emerged: ‘User Satisfaction’, ‘Cost-recovery Ratio’, and ‘Modal Split’. Concerning organisational features, the block of five higher rated features highlights consensus amongst consulted experts in relation to the importance of two key organisational dimensions, planning and funding, and a key organisational characteristic, integration. ‘Awarding Mechanism Employed’ and ‘Competition amongst Operating Companies’ have a less predominant role driving performance according to experts. The output from the GDPT enables proceeding with research efforts mentioned in Section 2.3. Nonetheless, its contribution is not restricted to this. Findings are rich, and the initial catalogue of observations deserves further exploration to yield new insights and inform policy interventions.

Figure 1: Profile of Experts



Describes experts that concluded Round 3: (a) experts' roles; (b) areas of technical expertise; (c) regional expertise. Respondents could choose multiple options – percentages refer to the total number of participants. 'Other' in Professional Role includes private think-thank, SIMUS (Latin American Association of Integrated Systems for Sustainable Urban Mobility), and Industry Association. 'Other' in Technical Expertise includes land use transport integration and costing & pricing.

ACKNOWLEDGEMENTS

Authors are grateful for the invaluable contribution of experts who participated in the survey and also thank colleagues Arman Noroozian, Christine Milchram, and Qasim Lone for their support.

REFERENCES

- Albalade, 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.
- 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.
- 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.
- 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.
- 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. (1987). *Managing Public Transit Strategically: A Comprehensive Approach to Strengthening Service and Monitoring Performance*. Jossey-Bass.
- Fielding, G. J. (1992). Transit performance evaluation in the U.S.A. *Transportation Research Part A*, 26A(6), 483–491.
- 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.
- Fiorio, C. V., Florio, M., & Perucca, G. (2013). User satisfaction and the organization of local public transport: Evidence from European cities. *Transport Policy*, 29, 209–218.
- Geerlings, H., Klementschtz, 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.
- Geurs, K., & van Wee, B. (2004). Accessibility evaluation of land-use and transport strategies: review and research directions. *Journal of Transport Geography*, 12, 127–140.
- 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.
- Grant, Y., Pollan, C., & Blake, T. (2011). *Regional Organizational Models for Public Transportation*. American Public Transportation Association.
- Hale, C. A. (2011). New approaches to strategic urban transport assessment. *Australian Planner*, 48(3), 173–182.
- Henkin, T., Padgett, R., Knapp, S., Fetting, E., Baker, S., & G, S. R. (2012). *TCRP Report 159: Transforming Public Transportation Institutional and Business Models*. Transit Cooperative Research Program.
- 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.
- 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*. (H. A. Linstone & M. Turoff, Eds.), *The Delphi Method - Techniques and Applications*. Addison-Wesley Educational Publishers Inc.
- 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.

- Mouwen, A., & Rietveld, P. (2013). Does competitive tendering improve customer satisfaction with public transport? A case study for the Netherlands. *Transportation Research Part A: Policy and Practice*, 51, 29–45.
- Ostrom, E. (1990). *Governing the Commons: The Evolution of Institutions for Collective Action*. New York: Cambridge University Press.
- Paredes-Molina, R., & Baytelman, Y. (1996). Urban Transport Deregulation: The Chilean Experience. *Estudios de Economía*, 23, 195–210.
- Preston, J., & Rajé, F. (2007). Accessibility, mobility and transport-related social exclusion. *Journal of Transport Geography*, 15, 151–160.
- Rihoux, B., & Lobe, B. (2009). The Case for Qualitative Comparative Analysis (QCA): Adding Leverage for Thick Cross-Case Comparison. In D. Byrne & C. C. Ragin (Eds.), *The SAGE Handbook of Case-Based Methods* (pp. 222–242). Sage Publications.
- 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 (JTEP)*, 41(2), 257–282.
- Scheffler, R., Hartwig, K. H., & Malina, R. (2013). The Effects of Ownership Structure, Competition, and Cross-Subsidisation on the Efficiency of Public Bus Transport: Empirical Evidence from Germany. *Journal of Transport Economics and Policy*, 47(Part 3), 371–386.
- 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 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., & Wallis, I. (2013). “Regulated deregulation” of local bus services-An appraisal of international developments. *Research in Transportation Economics*, 39(1), 21–33.
- Van Dijk, J. A. G. M. (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.
- 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.