Social Indicators Research
https://doi.org/10.1007/s11205-020-02284-0

ORIGINAL RESEARCH

An Integrated Approach to Select Key Quality Indicators in Transit Services

Benedetto Barabino[1] · Nicola Aldo Cabras[2] · Claudio Conversano[3] · Alessandro Olivo[4]

Accepted: 3 February 2020

Abstract

Recent interests in transit services have captured attention of experts on the monitoring of public transport quality. Previous research focused on relevant models and methods to monitor the quality of transit services and showed where and when different service quality levels occur. However, there was little attention to detect objectively a pool of key quality indicators (KQI) to be monitored, from a large set. This paper covers this gap by the proposal of an integrated approach, which identifies a long list of KQI, defines their properties, involves experts to elicit judgments for each KQI, evaluates the long list, and points out the most promising set. This integrated approach is demonstrated with an application based on an international survey and a Monte Carlo simulation method. Moreover, a restricted and relevant set of 9 overlapping KQI is derived by linking these results with those obtained from two different approaches.

Keywords Transit service quality monitoring · Transit service quality indicators · Monte Carlo simulation · Integrated approach

Mathematics Subject Classification C83 · L91 · L92 · R41 · R42

✉ Claudio Conversano
conversa@unica.it

Benedetto Barabino
benedetto.barabino@unibs.it

Nicola Aldo Cabras
nicolaaldocabras@gmail.com

Alessandro Olivo
olivo@unica.it

[1] Department of Civil Engineering, Architecture, Land, Environment and Mathematics (DICATAM), University of Brescia, via Branze 43, 25123 Brescia, Italy

[2] School of Economics, Law and Social Sciences, University of Cagliari, Viale S. Ignazio 78, 09123 Cagliari, Italy

[3] Department of Business and Economics Sciences, University of Cagliari, Viale S. Ignazio 17, 09123 Cagliari, Italy

[4] Department of Civil Engineering, Environment and Architecture, University of Cagliari, Via Marengo 2, 09123 Cagliari, Italy

Published online: 10 February 2020

1 Introduction

Over the past decade, Public Transport Companies (PTCs) were required to reduce costs and increase their performance mainly in terms of efficiency, effectiveness and quality (e.g., Hensher and Houghton 2004; Hensher and Stanley 2003). In this scenario, the service quality is measured through several indicators representing features of transit services. These indicators provide informative signals of excellences and criticalities in transit services, represent a strategic aspect of desired quality and are a fundamental input in the monitoring of quality. Thus, the selection of key quality indicators (KQI) is crucial for this monitoring. However, little attention has been paid on the following three issues.

First, earlier research has investigated about transit service performance by using a variety of key indicators. Nevertheless, in the specific domain of transit service quality, it has largely been focused on the development of models and methods orientated towards "how to measure and manage" KQI rather than on "what to measure" (de Oña and de Oña 2014). Moreover, it has seldom investigated indicator selection mechanisms using objective methods.

Second, usual efficiency and effectiveness indicators are a key to evaluate the performance of the transit service. Nevertheless, they are more focused on the PTC's viewpoint whilst the passengers' perspective is not always considered. However, this perspective is crucial, as passengers are the sole judges of service quality (Berry et al. 1990). Thus, KQI should be linked to the passenger perspective despite these KQI often return the PTC's viewpoint.[1]

Third, according to Castillo and Pitfield (2010), the selection of suitable KQI presents some challenges and requires a systematic method to improve the acceptability and credibility of KQI among experts. Many potential KQI may be considered, but the selection of a compact subset may be tricky. Moreover, since KQI are only constructs of the system, selecting those more suitable to characterize the system is challenging.

In view of the previous drawbacks and challenges, the objective of this paper is to propose a novel integrated approach, hereinafter IA, to identify and select a pool of KQI able to provide a high-level direction for the monitoring of transit service quality. IA identifies a long list of KQI, defines components and attributes for KQI, involves experts to elicit judgments, evaluates and adjusts KQI to account for coverage bias using Monte Carlo simulation methods and points out the most promising set. IA is evaluated on *ad-hoc* data gathered from an international survey involving leading academics and practitioners in the transit service quality.

The remaining of the paper is organized as follows: In Sect. 2, the related literature is reviewed. In Sect. 3, the IA is proposed to identify and select, in six steps, the most suitable KQI describing the quality of bus transit services on fixed routes. The implementation of the first two steps of IA is discussed in Sect. 4, whilst that of the third step, i.e., the survey, is presented in Sect. 5. Information about the implementation of the other three steps of IA is shown in Sect. 6, where results are presented, discussed and linked to previous research. Finally, conclusions and research perspectives are reported in Sect. 7.

[1] For instance, the regularity can be measured in terms of the percentage of buses which maintain evenness headways at bus stops—transit—perspective orientation—or in terms of the percentage of passengers who wait at bus stops less than a fraction of scheduled headways—passenger-perspective orientation (e.g., Barabino et al. 2017).

2 Literature Review

Table 1 lists the recent literature about indicators for the monitoring of transit services. It provides a summary of service aspects, indicators and sub-indicators based on the classification introduced by CEN/TC 320 (2002).[2] Moreover, Table 1 reports their classification according to the economic viewpoint, their domain (physical or operational), their type (hard or soft) and the way as these indicators have been selected for their monitoring. Items in Table 1 have been ordered according to the number of sub-indicators (i.e., column #sP).

It is worth noting that most studies did not report a classification based on three levels as in CEN/TC 320 (2002). However, to facilitate comparisons, we have analysed these studies with respect to these three levels.

Data reported in Table 1 leads us to the following considerations.

First, indicators (or sub-indicators) can be classified according to the economic viewpoint of the efficiency, effectiveness and quality of service. The efficiency directly focuses on the input to the produced service, whereas the effectiveness focuses on the input to the consumed service (Hensher 2007). In both cases, transit-oriented indicators are adopted for measurement (e.g., vehicle/km, cost/km, incidents/km, passengers/km, passengers/hour, revenue/hour, etc.). Conversely, the quality focuses on the output to the delivered service. Transit-oriented KQI (e.g., the percentage of on-time buses, the percentage of regular buses, the headway regularity) and passenger-oriented KQI (e.g., the waiting time at stops, the cleanliness, the space on board, the percentage of passengers receiving a punctual service) have been distinguished from each other.

Second, the domain of the indicators may be physical and/or operational (Barabino and Di Francesco 2016; Mahmoud et al. 2011). Key indicators evaluating transit service performance from a planning perspective characterize the physical domain.[3] The operational domain considers the set of key indicators used to evaluate service quality from the passengers' perspective, when passengers rate their satisfaction through specific indicators.

Third, hard and soft key indicators have been observed. The former may be viewed as quantitative indicators; the latter may be considered as qualitative ones. Thus, hard key indicators may be considered as more objective as opposed to soft ones, which can be influenced by personal opinions. In this last case, a careful definition of the key indicators helps derive a more objective assessment.

Fourth, the selection of suitable key indicators has been made by four approaches. The first approach selects key indicators by drawing on previous relevant studies about the monitoring of service quality, specifically norm-based guidelines, literature and/or explicit expertise in several economic and industrial fields. The second approach selects key indicators using focus groups and/or surveys on users, operators and/or experts. Using the third approach, indicators were selected according to key objectives of the organization e.g., running time adherence and headway regularity.

The fourth approach uses model-based (e.g., Eboli and Mazzulla 2015; Prioni and Hensher 2000) or algorithmic-based (e.g., Barabino and Di Francesco 2016; Mahmoud et al.

[2] CEN/TC 320 (2002) issued the European Norm EN 13816:2002 for the definition, targeting and measurement of transit service quality. See the "Appendix" for details.

[3] These indicators are usually evaluated by PTCs for the purposes of standardization, benchmarking and quality certification.

Table 1 The wide size of indicators and some properties. *Source*: Adapted and expanded from Barabino and Di Francesco (2016)

Source	#SA	#P	#sP	Economic viewpoint				Domain		Type		Indicator selection
				TE	Te	TQ	PQ	P	O	H	S	
Kittelsson et al. (2003b)[a]	10	32	191	•	•	•	•	•	•	•	•	L
Hirschhorn et al. (2018)	9	19	179	•	•	•	•	•	•	•	•	S
CEN/TC 320 (2002)[b]	8	30	106				•	•	•	•	•	L
Eboli and Mazzulla (2012)	8	20	51				•	•			•	L
MORPACE International et al. (1999)	8	18	48			•	•	•		•	•	S
Prioni and Hensher (2000)	8	18	45	•	•	•	•	•	•	•	•	*M*
Hensher (2014)	8	18	45	•	•	•	•	•	•	•	•	*S*
Hensher (2015)	8	18	45	•	•	•	•	•	•	•	•	*S*
Tyrinopoulos and Antoniou (2008a)	8	18	39			•	•	•	•	•	•	L
Eboli and Mazzulla (2015)	7	16	33			•	•		•	•	•	M
Friman et al. (2001)	7	15	33				•		•	•	•	L
Aydin et al. (2015)	*7*	*14*	*31*			•	•	•	•	•	•	*L*
Nathanail (2008)	8	14	30			•	•	•	•	•	•	L
Eboli and Mazzulla (2011)	7	14	26			•	•	•	•	•	•	L
Eboli and Mazzulla (2009)	7	14	26				•		•	•	•	L
Tyrinopoulos and Antoniou (2008b)	8	15	23			•	•		•	•	•	L
Parasuraman et al. (1991)	7	13	21				•		•		•	M
Guirao et al. (2016)	5	10	19	•		•	•	•	•	•	•	O
Barabino and Deiana (2013)	8	14	18				•		•		•	L
Sheth et al. (2007)	7	12	18	•	•	•	•	•		•		L
Eboli and Mazzulla (2007)	7	13	16				•		•	•	•	L
Eboli and Mazzulla (2010)	6	14	16			•	•		•	•	•	L
Stradling et al. (2007)	7	11	16				•		•	•	•	S
Barabino et al. (2012)	8	13	15				•		•	•	•	L
de Oña et al. (2016)	7	11	13				•	•	•	•	•	O
Hassan et al. (2013)	3	4	13	•	•		•	•	•	•	•	L
Moffat (2014)	8	11	13	•	•		•	•	•	•	•	L

Table 1 (continued)

Source	#SA	#P	#sP	Economic viewpoint				Domain		Type		Indicator selection
				TE	Te	TQ	PQ	P	O	H	S	
Badami and Haider (2007)	2	4	12	•	•			•		•		L
Too and Earl (2010)	6	11	11				•		•		•	L
Barabino et al. (2011)	5	8	10				•		•	•	•	L
Barabino et al. (2015a)	8	9	9				•	•	•	•	•	L
Beirão and Cabral (2007)	4	6	9				•		•	•	•	S
Ayadi and Hammami (2015)	1	3	8	•				•		•		L
Eboli et al. (2018)	5	7	7			•	•		•	•	•	L
Furth et al. (2006)	2	3	7			•	•	•		•		O
dell'Olio et al. (2011)	3	6	6				•	•		•	•	S
Hensher and Houghton (2004)	2	4	6	•				•		•	•	O
Kittelsson et al. (2003a)	3	5	6			•	•	•		•		L
Barabino and Di Francesco (2016)	2	3	3				•	•	•	•	•	M
Cantwell et al. (2009)	3	3	3			•	•		•	•		L
Chen et al. (2009)	1	1	3			•		•		•		O
Liekendael et al. (2006)	2	2	3				•	•		•	•	L
Lin et al. (2008)	1	2	2			•		•		•		O
Barabino et al. (2015b)	1	1	1				•	•		•		L

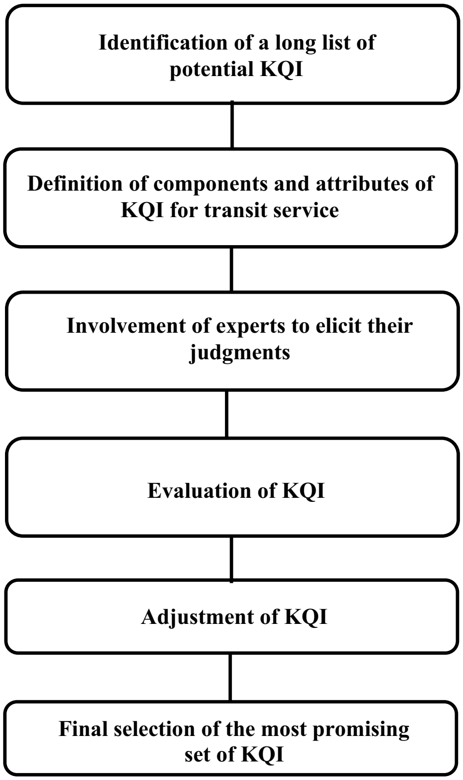
This is a representative but not a comprehensive list of references, which has been derived from the overall number of citations, authors' research and operational practice

#SA number of service aspect, #P number of indicators, #sP number of sub-indicators, TE transit-oriented efficiency, Te transit-oriented effectiveness, TQ transit-oriented quality, PQ passenger-oriented quality, P physical, O operational, H hard, S soft, L norm-based guidelines, literature and/or specific knowledge in some economic and industrial fields, M models and methods, O specific objectives of the organization, S surveys on users, operators and/or experts and/or focus groups

[a]This study contains a comprehensive list of 461 sub-indicators. However, 178 sub-indicators depending on the size of population have been disregarded as they contained a subset of the same 191 sub-indicators; 83 sub-indicators have been disregarded as they contained a subset of the same 191 sub-indicators, even if faced from a different perspective; 9 sub-indicators have been disregarded as they mainly referred to demand responsive systems

[b]This study contains a comprehensive list of 103 sub-indicators. However, 106 sub-indicators have been considered as 3 indicators have not sub-indicators associated. Therefore, we considered the sub-indicator as the indicator

Fig. 1 The proposed integrated approach for identifying and selecting relevant KQI



2011) methodologies for the selection of relevant indicators. For instance, Prioni and Hensher (2000) adopted discrete choice models that combined stated and revealed preference data. Conversely, Mahmoud et al. (2011) proposed a theoretical methodology based on two-steps: they derived a hierarchy of importance and a set of desired indicators from a panel of experts and a sample of passengers using both qualitative and quantitative methods. Next, they defined a concise set of key indicators by merging the outcomes of the first step. However, this method was not applied in a real context.

Our paper focuses on the fourth approach. More precisely, IA builds on the first block of TRANSQUAL recently introduced in Barabino and Di Francesco (2016). They pointed out some guidelines to drive in the selection of KQI and suggested the use of an effective methodology when the merging between KQI' list provided by users and experts results in "close to zero". Moreover, IA differs from Evaluative and Logical Approach to Sustainable Transport Indicator Compilation (ELASTIC), which is a framework for the identification and selection of sustainable transport indicators by UK experts (Castillo and Pitfield 2010). In addition, it differs from Eboli and Mazzulla (2015) both in the method and in the field of application.[4]

[4] Eboli and Mazzulla (2015) began the analysis by choosing among 33 sub-indicators and restrict the investigation to a railway transit system.

3 Methodology

To identify and select a pool of KQI,[5] our integrated approach (IA) is composed of six steps represented in the scheme of Fig. 1.

Step 1: Identification of a long list of potential KQI In this step, one identifies the relevant literature according to the academic and practical viewpoints and derives the more comprehensive initial source of potential KQI.

Step 2: Definition of components and attributes of KQI In deciding which KQI should be selected, two manageable components are firstly considered:

- The methodological features of KQI.
- The relevance of KQI to the concept of service quality.

These components recognize some high-level characteristics of KQI. However, these components could be too vast to drive in the selection of KQI. As a result, an additional decomposition into more manageable (i.e., measurable and accurate) attributes is required. The literature suggests several methodological attributes of desirable key indicators (e.g., Castillo and Pitfield 2010; Dhakal and Imura 2003). Drawing on these studies, four attributes are identified.

- *Measurability* It should be possible to measure indicators in a theoretical sound and dependable manner.
- *Ease of availability* It should be possible to collect data at a reasonable cost in line with the requirements of parsimonious PTCs.
- *Speed of availability* It should be possible to update the assessment regularly, so that the time between consecutive measurements can be reduced.
- *Interpretability* It should be possible to consider indicators unambiguous and universally understandable. Indicators expressed in numerical values such as percentages or letters are preferred, as these outputs are more user-friendly for technicians, senior managers and users.

These attributes are general in the application of the method. However, the level of importance assigned to each of them can vary to reflect different viewpoints.

Furthermore, four main attributes related to the quality components are considered. These attributes are obtained from Barabino and Di Francesco (2016) who interpreted the service quality norms EN 13816:2002 (CEN/TC 320 2002) and EN 15140:2006 (CEN/TC 320 2006)[6] considering the relationship between indicators introduced by

[5] For the sake of synthesis, in what follows, service aspects, indicators and sub-indicators will all be referred to as indicators or KQI.

[6] EN 13816:2002 is a ground-breaking standard designed to enhance the promotion of a more customer-oriented quality approach within the public transportation sector. It is intended to be adopted by PTCs in the presentation and monitoring of their services, but it is recommended for use by authorities and transit agencies for the procurement of public passenger transport services in the preparation of invitations to tender.

EN 15140:2006 is intended to help construct the measurement system and to help understand and reduce the causes of biases that any system of measurement may introduce. Moreover, a set of reference levels needed to measure the degree of fulfilment of the EN 13816:2002 indicators has been introduced, so to help in the adoption of standardized measurement methods and operational procedures for the service quality determination.

the regulation and service quality. Therefore, these attributes are expected to contribute towards the acceptability of indicators among practitioners also. They are:

- *Integration between users' and PTC's perspectives* The evaluation of the quality using the indicator should re-orient the organisation as more customer-focused, i.e.: the PTC should consider the user as the key component instead of a generic and, sometimes, unimportant component.
- *User-orientation* This is a challenging aspect of overall quality assessment, as when there are several indicators of a predefined aspect of quality, it is important to avoid PTC-oriented parameters that cannot be understood by users. For instance, the measure of safety in terms of the presence of usable handrail/handhold versus injured/km.
- *Measurability from subjective and objective perspectives* Both subjective and objective data should contribute equally to the quality monitoring process to ensure uniformity. The requirement to integrate subjective and objective indicators obliges the PTC to evaluate the impact of performance on user satisfaction for each indicator and to collect evidence about the passenger perceptions on delivered service.
- *Quantification of the number of passengers* Indicators should include a measure accounting for the number of passengers. For instance, when automatic passenger counting system is adopted by a PTC, it is not difficult to evaluate passenger numbers, but there are other sources of passenger data such as sold tickets and electronic farebox solution. In the case of both derived and estimated passenger data, the use of percentages is recommended to guarantee the same unit of measure. Although this approach might disregard possible heterogeneity among locations related to differences in country developments and reference to spatial units, it results the most consistent with the norms. According to the EN 13816:2002 and EN 15140: 2006, the level of achievement of the objective shall be expressed, where possible, as a ratio of passengers affected: this kind of output helps prioritize actions among competitive routes of the same network owing to the quantification of the number of passengers affected (CEN/TC 320 2002, 2006).

The level of importance assigned to each quality related attribute can vary as well as for methodological attributes. Moreover, these attributes are expected to derive KQI considered as SMART, i.e., Specific, Measurable, Achievable, Realistic and Timely (Doran 1981).

Step 3: Involvement of experts to elicit their judgments Academics and practitioners are involved to drive the selection of KQI. Academics have a profound knowledge of the meaning of each indicator and its properties at high level; thus, they are expected to provide an accurate and theoretically sound evaluation of components and attributes. At the same time, practitioners provide useful information about the viability of components and attributes for the indicators at hand, because they manage daily the quality to keep passengers and attract new ones. In what follows, academics and practitioners are referred to as "experts".

Several approaches can be used for the involvement of experts, even if there is no single way to operate. The proposed IA is based on an international survey to possibly involve many experts. In reflecting on the type of survey, the choice of a web-survey is suggested because of some advantages, which make this procedure also well practiced elsewhere (e.g., Too and Earl 2010).

In the proposed survey, experts are involved twice.

First, they help derive the importance of components and attributes as the focus is on the perceived value of each component and attribute by experts. Since this perception can vary due to the specific knowledge on KQI, and, thus, provide biased opinions towards components and attributes, a weighing process may help overcome this problem. Hence, weights of importance are attached to components and attributes (*or* items). According to e.g., Zelany (1974) and Castillo and Pitfield (2010), even if weights can be directly attached by questioning experts on preferences for the single item, this approach may be flawed since humans have difficulties in processing relevant information into stable weights when too many items exist. Moreover, even if e.g., Cantwell et al. (2009), CEN/TC 320 (2006), Eboli and Mazzulla (2011), Friman et al. (2001), Hassan et al. (2013), Wang and Lee (2009) proposed several approaches to weighting items, IA utilises the Analytical Hierarchy Process (AHP). AHP establishes weights from a pair-wise comparison of items and generates a ratio scale from each set of comparisons to deal with inconsistency in judgments (Saaty 1980, 1987, 1990).[7] AHP is applied twice: first, experts compare the two components, the methodological features of KQI and the relevance of KQI to the concept of service quality and, next, the two sets of four attributes characterizing each component.

Second, experts rate each KQI against each attribute according to predefined scales (e.g., numerical, qualitative and so on).

Step 4: Evaluation of KQI In this step, the score of each KQI is computed. This score is defined as *Weighted Key Quality Indicator Score* (*WKQIS*) and is obtained from simple additive weighting. The latter aggregates weights and outcome marks to evaluate the performance of each KQI. More precisely, let:

- I be the set of KQI;
- J be the set of experts involved;
- H be the set of the methodological attributes;
- K be the set of the quality attributes;
- m_j be the weight of the methodological component as derived by AHP, according to the evaluation of expert $j \in J$;
- q_j be the weight of the quality component as derived by AHP, according to the evaluation of expert $j \in J$;
- w_{jh} be the weight of attribute $h \in H$ as derived by AHP, according to the evaluation of expert $j \in J$;
- w_{jk} be the weight of attribute $k \in K$ as derived by AHP, according to the evaluation of expert $j \in J$;
- $\bar{M}$ be the average weight of the methodological component;
- $\bar{Q}$ be the average weight of the quality component;
- $\overline{w_h}$ be the average weight of attribute $h \in H$;
- $\overline{w_k}$ be the average weight of attribute $k \in K$;
- V_{ijh} be the mark of indicator $i \in I$ for attribute $h \in H$ according to the evaluation of expert $j \in J$;

[7] Although this method may be questionable (Dyer 1990), many scholars defend it and document about the advantages deriving from its use (e.g., Saaty 1990; Harker and Vargas 1990; Forman and Gass 2001; Ramanathan 2001; Millet and Wedley 2002; Macharis et al. 2004; Oguztimur 2011). Moreover, AHP is widely adopted in many fields of engineering, which contribute to reinforce its use among bus operators (e.g., de Steiguer et al. 2003).

- V_{ijk} be the mark of indicator $i \in I$ for attribute $k \in K$ according to the evaluation of expert $j \in J$;
- $\underline{V_{ih}}$ be the average mark of indicator $i \in I$ for attribute $h \in H$;
- V_{ik} be the average mark of indicator $i \in I$ for attribute $k \in H$.

For each indicator i, the *WKQIS* is computed according to the following four-step algorithm.

1. Compute $\bar{M}$ and $\bar{Q}$

$$\bar{M} = \frac{\sum_{j=1}^{J} m_j}{J} \tag{1}$$

$$\bar{Q} = \frac{\sum_{j=1}^{J} q_j}{J} \tag{2}$$

2. Compute $\overline{w_h}$ and $\overline{w_k}$

$$\overline{w_h} = \frac{\sum_{j=1}^{J} w_{jh}}{J} \quad \forall h = 1, \dots, \bar{H} \tag{3}$$

$$\overline{w_k} = \frac{\sum_{j=1}^{J} w_{jk}}{J} \quad \forall k = \bar{H} + 1, \dots, \bar{K} \tag{4}$$

3. Compute $\overline{V_{ih}}$ and $\overline{V_{ik}}$

$$\overline{V_{ih}} = \sum_{j=1}^{J} \frac{V_{ijh}}{J} \quad \forall h = 1, \dots, \bar{H}; \quad \forall i = 1, \dots, \bar{I} \tag{5}$$

$$\overline{V_{ik}} = \sum_{j=1}^{J} \frac{V_{ijk}}{J} \forall k = \bar{H} + 1, \dots, \bar{K}; \forall i = 1, \dots, \bar{I} \tag{6}$$

4. Compute $WKQIS_i$

$$WKQIS_i = \bar{M} \cdot \left(\sum_{h=1}^{\bar{H}} \overline{w_h} \cdot \overline{V_{ih}} \right) + \bar{Q} \cdot \left(\sum_{k=\bar{H}+1}^{\bar{K}} \overline{w_k} \cdot \overline{V_{ik}} \right) \quad \forall i = 1, \dots, \bar{I} \tag{7}$$

Step 5: Adjustment of KQI The adjustment of the $WKQIS_i$ is needed to reduce or eliminate coverage bias. The latter negatively affects findings obtained from web-based surveys when the number of the responses returned by experts is not too large (e.g., Dever et al. 2008). In IA, the adjustment of $WKQIS_i$ is done by Monte Carlo simulation. The deterministic model of judgements is iteratively evaluated using a set of random numbers as input. Hence, the observed (deterministic) model is turned into a stochastic model.

Usually, the theoretical model for deriving $WKQIS_i$ is simulated in order to produce artificial data that need to be matched with observed data (i.e., judgements by experts).

Matching is obtained by choosing a suitable probability distribution that best represents the current state of knowledge. Specifically, to derive an adjusted set of $WKQIS_i$, Monte Carlo simulation is applied twice to retrace the whole process designed in the theoretical model leading to the computation of $WKQIS_i$ according to Eq. (7). Thus, two experiments are designed, namely:

- *Experiment 1* the distribution of the scores obtained for the entire set of $|H|+|K|$ items characterizing AHP is simulated B times by multinomial distributions with $|H|+|K|$ classes. The choice of the multinomial distribution is motivated by the consideration that $|H|+|K|>2$, in most cases. Thus, a multi-class reference distribution is required. Of course, for the case $|H|+|K|=2$ the shift to a binomial distribution is straightforward. Scores obtained from pairwise comparisons of items made by experts are used as a priori probability of the $|H|+|K|$ classes in the Data Generating Process (DGP). DGP is based on several trials (sample size) corresponding to the number of participants involved in the survey and is simulated B times. The final adjusted weights for each component and attribute are the average of mean values obtained in each trial.
- *Experiment 2* the same process used in Experiment 1 is repeated for the set of KQI related to the $|H|+|K|$ attributes. Usually, set I is very large and the total number of items considered in the simulation is $S=(|H|+|K|)\times|I|$. For each indicator $i\in I$, the distribution of scores obtained by the experts is simulated B times using a binomial distribution X defined in the [1,10] interval, such that $E(X)=\overline{V_{ih}}/|J|$ or $E(X)=\overline{V_{ik}}/|J|$ if indicator $i\in I$ concerns an attribute of a methodological or a quality component, respectively. More precisely, the mean score obtained for indicator $i\in I$ is rescaled on the [0,1] interval and is used as the indicator of the Binomial distribution simulated for a number of trials corresponding to the number of items characterizing a specific indicator. Again, the average score obtained for each indicator derives from averaging scores obtained in the simulation trials.

Step 6: Final selection of KQI In the last step of the IA, KQI are ranked in decreasing order according to their estimated $WKQIS_i$. The best indicator $i^*\in I$ presents the highest estimated $WKQIS_i$. More precisely, a double ordering of KQI is defined. In the former, $WKQIS_i$ are recomputed by replacing empirical weights with those obtained from Experiment 1, and recomputed $WKQIS_i$ are ordered decreasingly. In the latter, $WKQIS_i$ are recomputed by replacing empirical marks with those of Experiment 2. The final rank assigned to each individual $WKQIS_i$ is the average rank obtained from the double ordering. The output of the process is the "adjusted" selection of a pool of top n KQI as representative of the quality of transit services on fixed routes.

4 The Long List of KQI

According to Step 1 of IA, the more comprehensive initial source of potential KQI is first retrieved. As shown in Table 1, Kittelson et al. (2003b), Hirschhorn et al. (2018) and CEN/TC 320 (2002) can be considered the main sources for this purpose. Indeed, these three references contain all the indicators included in the other studies, even if some KQI present a slightly different name.[8]Nevertheless, even if Kittelson et al. (2003b) and Hirschhorn et al. (2018) provided the largest list of indicators, the initial list was retrieved by CEN/TC 320 (2002). This choice is motivated as follows.

- It does not include efficiency and effectiveness indicators, but only quality ones.
- It considers the most relevant long list of available service aspects (8) and indicators (30) as compared to Kittelson et al. (2003b) where 8 service aspects and 24 indicators are included, respectively. Conversely, Kittelson et al. (2003b) include more sub-indicators than CEN/TC 320 (2002) (i.e., 135 vs 106). This is because CEN/TC 320 (2002) adopts only sub-indicators, whilst Kittelson et al. (2003b) add further and related sub–sub-indicators, perhaps to provide a detailed comprehension. For instance, CEN/TC 320 (2002) considers the sub-indicator cleanliness only and points out that further and related sub–sub-indicators may be specified by the PTC. Kittelson et al. (2003b) specify some sub–sub-indicators for the sub-indicator cleanliness, such as windows cleanliness and vehicle interior cleanliness. However, these differences do not influence our choice, because all sub–sub-indicators may be clustered into the same sub-indicator. In addition, despite Hirschhorn et al. (2018) have more service aspects and sub-indicators, CEN/TC 320 (2002) have more service aspects (8 vs 7) and sub-indicators related to service quality only (106 vs 53).
- It emphasises the use of user oriented and measurable KQI for quality monitoring, even if it is not univocally specified how to choose a relevant sub-set of them.
- It represents a primary tool among worldwide PTCs to monitor and certify the quality of service of their routes (e.g., AFNOR 2005a, b; Barabino et al. 2013; Barabino 2018).

Components and attributes pointed out in Step 2 of IA are the input data of the selection process.

5 The Survey

Next, according to Step 3 of IA, experts are involved to elicit their judgments on components, attributes and marks for each indicator by a web-based international survey. More precisely, because we aim to find a restricted set of KQI to be used by any organisation, experts of all continents were involved to form a heterogeneous panel.

Two different approaches have been adopted for selecting experts: one for academics and one for practitioners. Academics were selected from both a recent survey on the top public transportation scientists (Heilig and Voß 2015) and the ranking of the top 50 worldwide universities classified according to the Center for World University Rankings[9] and to the Academic Ranking of World Universities,[10] in 2017. Moreover, the Times Higher Education ranking[11] has been considered to gather the information about the top universities on transportation. Based on these criteria, 291 e-mail addresses have been selected from the websites of the universities. Since no available ranking was detected for PTCs' worldwide,

[8] For instance, the waiting time is not explicitly mentioned in CEN/TC 320 (2002), but it can be derived easily from measurements of regularity and/or punctuality (e.g., Barabino et al. 2017).

[9] http://cwur.org/2016.php.

[10] http://www.shanghairanking.com/.

[11] https://www.timeshighereducation.com/.

Table 2 PTCs' distribution worldwide. *Source*: Authors' elaboration from data of PTCs' UITP member

Continent	Companies	Distribution (%)	No. of selected Public Transport Companies
Europe	435	70	35
North and South America	99	16	8
Africa	61	10	5
Asia–Pacific	14	2	1
Australia and New Zealand	16	2	1
	615	100	50

a different approach was used to select a panel of practitioners. The *Union International des Transports Publics* (UITP)[12] represented the main source. UITP provided a list of 615 PTCs' members, from which 50 PTCs were selected according to their geographical distribution (see Table 2). It is worth noting that the selection of the PTCs' members followed a random sampling criterion stratified by continent.

The personal contact of PTCs' experts was trickier to find than that of academics, as multiple sources have been adopted. First, search engines like google.com were used to visit the websites of PTCs. Next, LinkedIn[13] was adopted to search several profiles of PTCs' experts in order to retrieve their specific skills. In cases where no emails were found, the search was refined using Skrapp.io[14] inside LinkedIn. The combined use of LinkedIn and Skrapp.io allowed us to obtain the email of the practitioners to contact. They were selected based on their position in the company. Furthermore, GrowthBot[15] was used to check basic information about company as the address and the name of the competitors in order to identify new members to be included in the survey.

Nevertheless, to preserve homogeneity of respondents with respect to academics and practitioners, a final list of 214 sample units of 107 academics and 107 practitioners was selected from the original list of 604 email addresses based on a random sampling scheme.

The survey was organized in Stage 1 and Stage 2 and required three waves of data collection that were carried out between January 2018 and February 2019. A total of 214 e-mails presenting the aim and scopes of the research project and the link to the web-survey were sent. Since the survey has been addressed to experts, it was supposed that questions were not difficult to understand. Nonetheless, to prevent possible misunderstandings, each e-mail contained a link to retrieve the definitions of component and attributes of each indicator.

Stage 1 was carried out in the first wave. In this stage, experts were required to perform pairwise comparisons of both methodological and relevance of quality components, as well as related attributes. Their responses were evaluated on a 1–9 points scale (see Table 3), which is an adjusted version of the fundamental scale of Saaty (1980; 1987). Questions concerning the first and the second pairwise comparisons were formulated in order to collect data in a way amenable for performing AHP and are shown

12 http://www.uitp.org/. UITP is a no profit international organization including 1,400 members from 96 countries worldwide.

13 https://it.linkedin.com/.

14 www.skrapp.io.

15 www.growthbot.org.

Table 3 Scale to rate components and attributes

Level of importance	Definition	Comment
1	Equal importance	Two items (i.e., components and/or attributes) contribute equally
3	Moderate importance	Experience and judgment slightly favour one item over another
5	Essential or strong importance	Experience and judgment strongly favour one item over another
7	Very strong importance	An item is strongly favoured and its dominance demonstrated in practice
9	Extreme importance	The evidence of one item over another is of the highest possible order of confirmation
2, 4, 6, 8	Intermediate values between the two adjacent	There might be cases when experience and judgment may not make an item comparable to another according to the five points scale. Therefore a "in between" value may be more suitable

If you were asked to choose a transit service quality parameter based on either its Methodological Feature or its Relevance to service quality which of the two criteria would you deem more important to your selection and how strongly so?

Please mark the appropiate number (1 = equal importance)

	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	
Methodological Quality																		Relevance to Quality

In deciding a parameter of 'Transit Service Quality', which of the following attributes would you deem more important for indicator choice and how strongly?

Please mark the appropiate number (1 = equal importance)

	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	
Measurability																		Ease of Availability
Speed of Availability																		Measurability
Interpretability																		Measurability
Speed of Availability																		Ease of Availability
Ease of Availability																		Interpretability
Interpretability																		Speed of Availability
Integration of users' and companies' views																		User Orientation
Subjective & Objective Measurability																		Integration of users' and companies' views
Expressing the amount of passengers																		Integration of users' and companies' views
User Orientation																		Subjective & Objective Measurability
User Orientation																		Expressing the amount of passengers
Subjective & Objective Measurability																		Expressing the amount of passengers

Fig. 2 Stage 1. Example of questionnaire for pair-wise comparisons

Considering a range from 1 (the worst) - 10 (the best) please give a mark for every parameter of **Availability** - **Network** of Public Transport

	Measurability	Ease of Availability	Speed of Availability	Interpretability	Integration Users-Companies	User Orientation	Subjective & Objective Measurability	Amount of Passengers
Distance to b/a-points	1-10	1-10	1-10	1-10	1-10	1-10	1-10	1-10
Need to transfer	1-10	1-10	1-10	1-10	1-10	1-10	1-10	1-10
Area Covered	1-10	1-10	1-10	1-10	1-10	1-10	1-10	1-10

Fig. 3 Stage 2. Example of questionnaire for the evaluation of KQI

in Fig. 2. For the sake of simplicity, methodological attributes and quality ones are evaluated in the same questionnaire considering pairwise comparison for each group. Besides, first experts compare the two components; next, they compare the four methodological attributes and the four quality ones. The first wave of Stage 1 was completed by 53 experts (24.73% of the panel). Collected data have a global consistency ratio (CR) of 0.80%.

Stage 2 was carried out during the second and third waves of the survey. Experts were required to grade each indicator against each attribute and to assign it a specific mark defined on a 1 (the worst) to 10 (the best) scale. The choice of this scale does not affect the generality of the method. Although more response options lead to somewhat lower scores, Dawes (2008) pointed out the similar reliability of different scales from a statistical viewpoint. Figure 3 shows an example taken from the questionnaire related to the service aspect 'Availability', the indicator 'Network', and the sub-indicators: 'distance to b/a points', 'need for transfer' and 'area covered'. This marking was required for all the 106 considered sub-indicators. The second and third waves of the survey in Stage 2 were completed by 43 experts, which represent about 20% of the total panel.

Table 4 Sample composition and participation in the survey

Experts	Continent	E-mail addresses randomly selected and sent [#]	Wave 1		Waves 2 and 3	
			Replies [#]	Response Rate	Replies [#]	Response Rate
Academics (291 e-mail addresses retrieved)	Asia–Pacific	11	5	0.05	4	0.04
	Australia and New Zealand	20	5	0.05	3	0.03
	Europe	40	8	0.07	7	0.07
	North and South America	36	13	0.12	12	0.11
	Total	107	31	0.29	26	0.24
Practitioners (313 e-mail addresses retrieved)	Asia–Pacific	9	2	0.02	1	0.01
	Australia and New Zealand	18	2	0.02	1	0.01
	Europe	46	16	0.15	14	0.13
	North and South America	34	2	0.02	1	0.01
	Total	107	22	0.21	17	0.16
	Total	214	53	0.25	43	0.20

Table 5 Weights of components

Component	Symbol	Mean weight	Standard deviation	Coefficient of variation (%)
Methodological features	$\bar{M}$	0.368	0.269	73.1
Relevance to transit service quality	$\bar{Q}$	0.632	0.269	42.6

Table 6 Weights of attributes

Attribute	Symbol	Mean weight	Standard deviation	Coefficient of variation (%)
Measurability	$\overline{w_1}$	0.2605	0.1769	67.9
Ease of availability	$\overline{w_2}$	0.2854	0.1559	54.6
Speed of availability	$\overline{w_3}$	0.1461	0.1201	82.2
Interpretability	$\overline{w_4}$	0.3079	0.1745	56.7
Integration users-companies	$\overline{w_5}$	0.1968	0.0981	49.8
User orientation	$\overline{w_6}$	0.3597	0.2072	57.6
Subjective and objective measurability	$\overline{w_7}$	0.2482	0.1698	68.4
Amount of passengers	$\overline{w_8}$	0.1953	0.1552	79.5

A summary of the information about the sampling design and the participation in the survey is reported in Table 4.

Overall, the participation of experts confirmed the interest in this topic and their evaluations can be considered as the first step in driving towards a preliminary choice of KQI. Interestingly, European practitioners returned the largest numbers of complete questionnaires, perhaps owing to the strongest interest in the EN 13816:2002 as opposed to practitioners of other continents. Besides, the response rate of the survey is consistent with Sivo et al. (2006), who investigate about the response rates reported in web-based surveys whose results have been published in six major journals on information systems.

6 Empirical Evidence

6.1 Weights, Scoring and Selection of KQI

Table 5 reports results about the average weights of the components, their standard deviation and coefficients of variation. Table 5 is self-explanatory. Nevertheless, it provides evidence that experts give more importance, on average, to the quality component than to the methodological one as the mean weight of $\bar{Q}$ is almost doubled with respect to $\bar{M}$. The *t* test about the difference in mean weights attributed by experts to the two components indicates that the observed difference is significant (*p* value <0.001). The superior importance of the quality component is probably due to the immediate impact that the first set of attributes has on the vision of quality intended by experts and is also enforced by the distribution of the scores of the two sets of attributes having the same standard deviation. Besides, the coefficient of variation in Table 5 provides evidence that the distribution of scores concerning the methodological features is much more dispersed than that of the other set of attributes.[16]

Furthermore, Table 6 reports results of the average weights and dispersion for each methodological (i.e., $h \in H$) and quality attribute (i.e., $k \in K$), respectively. Table 6 is self-explanatory: the weights show the differences obtained from experts' opinions with respect to the methodological and quality attributes.[17]

On the one hand, these outcomes show that experts identify 'Interpretability' and 'Ease of Availability' as the most important attributes to consider when evaluating the methodological features. These results show, as expected, that understandable and 'cheap' KQI can drive towards a sustainable and manageable system. On the other hand, 'User orientation' and 'Subjective & Objective Measurability' represent the most important attributes when evaluating the relevance to transit service quality. These results were also expected. Indeed, the User Orientation attribute allows PTCs to see its service as a user do. In addition, the Subjective & Objective Measurability attribute obliges the PTC to evaluate the impact of performance on the user satisfaction.

Next, according to Step 4 of IA, Original $WKQIS_i$ (O_$WKQIS_i$) are computed by Eq. (7). Experiments 1 and 2 described in Step 5 of IA have been performed with $B = 10{,}000$ and $S = 824$. For each indicator, the $WKQIS_i$ has been obtained and ranked in decreasing order for both original and adjusted values, according to Step 6 of IA. Results are shown in Tables 7 and 8, which report the list of top 26 KQI with the highest O_$WKQIS_i$ and Adjusted $WKQIS_i$, (A_$WKQIS_i$) respectively. As for the latter, we performed Experiment 1 described in Step 5 of IA mimicking $B = 10{,}000$ times the behavior of respondents w.r.t. the assignment of a weight to the two categories of KQI, and thus to the four attributes of the methodological features of KQI and to the four attributes concerning the relevance of KQI to the concept of service quality. The vector of observed weights was used to define the input parameters of the multinomial distribution and a new distribution of weights was simulated in each run. The average values of these weights over the B runs are the adjusted weights obtained for both the methodological and relevance components of KQI. Multiplying these estimated weights by the actual scores assigned by respondents to each KQI allows us to get a first set of adjusted KQI. Next, Experiment 2 described in Step 5 of IA was performed mimicking $B = 10{,}000$ times the behavior of respondents w.r.t. the assignment of a score to the (large) set of items characterizing the two categories of KQI. We simulated a scenario where, in each run, the respondent provided a score ranging from 1 to 10 for the eight attributes related to the two categories of KQI. The distribution of these scores was simulated based on the actual average score received by each item. Multiplying these estimated average scores by the actual weights assigned by respondents to each KQI allowed us to get a second set of adjusted KQI. These two sets of adjusted KQI were not ranked equally. Thus, as specified in Sect. 3 (Step 6), a weighted rank based on the double ordering of the two sets was considered to compute the A_$WKQIS_i$ represented in Table 8.

We focused our analysis on the restricted set of the first 26 KQI corresponding to the first quartile of the rank-ordering distribution of all indicators. Columns from 2 to 4 of

[16] For sake of completeness, the distribution of weights attributed to the quality and methodological components by academic and practitioners w.r.t. their geographical localization is reported in Table 10 in the "Appendix".

[17] For sake of completeness, the distribution of weights attributed to the attributes of quality and methodological components by academic and practitioners w.r.t. their geographical localization is reported in Tables 11 and 12 in the "Appendix".

Table 7 Top 26 KQI to monitoring the quality in transit service ranked according to O_WKQIS_i

Rank	Service aspect	Indicator	Sub-indicator	Trip stage	Mean score	SD
1	Availability	Operation	Frequency	Pre-trip	6.838	1.751
2	Time	Adherence to schedule	Regularity	En route	6.821	1.803
3	Information	General information	About security	In vehicle	6.812	1.623
4	Information	Travel information normal conditions	About fare	Pre-trip	6.788	1.458
5	Time	Adherence to schedule	Punctuality	En route	6.785	1.767
6	Information	Travel information normal conditions	About time	Pre-trip/en route	6.717	1.640
7	Information	Travel information normal conditions	About route	Pre-trip/en route	6.695	1.507
8	Comfort	Seating and personal space	At B/A points	En route	6.665	1.418
9	Comfort	Seating and personal space	In vehicle	In vehicle	6.649	1.599
10	Accessibility	Ticketing availability	Validation	En route	6.583	1.715
11	Customer care	Ticketing options	Payment options	En routen/in vehicle	6.552	1.606
12	Accessibility	Ticketing availability	Acquisition on network	Pre-trip/in vehicle	6.524	1.945
13	Information	Travel information normal conditions	About type of ticket	Pre-trip	6.510	1.328
14	Information	General information	About sources of information	Pre-trip/en route	6.491	1.802
15	Comfort	Ambient conditions	Noise	In vehicle	6.480	1.579
16	Information	General information	*About comfort*	Pre-trip	6.478	1.532
17	Information	General information	About travelling time	Pre-trip/en route	6.477	1.533
18	Comfort	Ambient conditions	Cleanliness	En routen/In vehicle	6.472	1.953
19	Information	General information	About accessibility	Pre-trip	6.470	1.883
20	Information	General information	*About availability*	Pre-trip	6.427	1.881
21	Time	Length of trip time	In vehicle	In vehicle	6.419	1.407
22	Availability	Operation	Vehicle load factor	In vehicle	6.411	1.689
23	Availability	Dependability	*Dependability*	Pre-trip/en route	6.392	1.420
24	customer care	Staff	Appearance	In vehicle	6.388	1.651
25	Security	Freedom from crime	Staff/police presence	Pre-trip/en route	6.369	1.278
26	Security	Freedom from accident	*Avoidance/visibility of hazards*	Pre-trip/en route	6.303	1.915

Tables 7 and 8 report the classification of Service Aspect, Indicator and Sub-Indicator recommended by CEN/TC 320 (2002),[18] to consider a well-known standard adopted by many European PTCs. Columns 5 to 7 report the trip stage, the scores and the standard deviations (SD) of O_$WKQIS_i$ or A_$WKQIS_i$, respectively.

Tables 7 and 8 provide interesting results.

First, the two sets are very similar and they share 22 of the 26 indicators. The four different indicators between the two sets are italic edited. The Spearman's rank correlation coefficients between the two sets of the 22 common indicators equals 0.77 (p value <0.001).

Second, seven of eight service aspects of CEN TC 320 (2002) are represented in this list by one or more KQI. Conversely, no KQI is included in the service aspect Environment. Thus, the findings suggest there is a minor attention towards it and a major attention toward classical KQI in the use of the transit service. Although this result could be unexpected, it might be justified as follows. The survey involved only academics and practitioners who are directly engaged in the evaluation of transit quality, whereas it has disregarded further experts perhaps more interested in this 'external' aspect. Moreover, it might be difficult to find fully user-oriented indicators reflecting this service aspect.

Third, all the trip stages are represented with a major attention to the pre-trip KQI (15 indicators of 36 and 14 of 36 as shown in Tables 7, 8, respectively). Less attention is given to the en route KQI (11/36 and 11/36) and to the in-vehicle KQI that result as the less-represented trip stages (10/36 and 11/36). This biased distribution towards the pre-trip-stage seems depend on service aspect Information as it includes about half of the top 26 KQI.

Fourth, these results show that adjustment of original scores partially changed the ranking emerged by experts' evaluations (*O_*$WKQIS_i$). Moreover, some relevant changes on the score of several KQI appear when comparing original scores with adjusted ones. It is noteworthy that adjustment tends to reduce *O_*$WKQIS_i$ in most cases. This may depend from the distribution of the scores obtained for a certain item from all experts. Indeed, the distribution is rather homogeneous with a tendency towards central values included in the interval from 5 to 7, except for some sporadic high scores whose effect on the mean score is mitigated by Monte Carlo simulations. Moreover, for KQI 'About comfort' and 'About availability', 'Dependability' and 'Avoidance/Visibility of Hazards' edited in italics in Table 7, the adjustment worsens consistently their ranking: they are not included in the list of top 26 adjusted KQI. Conversely, for KQI 'Modes', 'For customers needing help', at b/a points' and 'Toilets/Washing' edited in italic in Table 8, the adjustment consistently improves their ranking. Indeed, these indicators are now included in the top 26 adjusted KQI, even if they were in a lowest position in the original ones. For these cases, the original distribution of original scores is very homogeneous so that adjustment increases their scores.

Last but not least, the adjustment provides evidence that the three best indicators concern 'Mode', 'Regularity' and 'Information on time', instead of 'Frequency', 'Regularity' and 'Information on time' derived from original $WQKIS_i$. Although the 'Frequency' is not included in the adjusted $WQKIS_i$, the results are partially in accordance with Balcome et al. (2004). Indeed, they reported that service reliability (that includes time KQI) is twice as important as frequency and almost seven times more important than information for passengers. Besides, even if 'Frequency' is not ranked in the top three positions, it is included in the list of the best 26 KQI.

[18] See Table 9 in the "Appendix" for details.

Table 8 Top 26 KQI to monitoring the quality in transit service ranked according to A_WKQIS_i

Rank	Service aspect	Indicator	Sub-indicator	Trip stage	Mean score	SD
1	Availability	Modes	*Modes*	Pre-trip/en route	6.494	1.234
2	Time	Adherence to schedule	Regularity	En route	6.596	1.229
3	Information	Travel information normal conditions	About time	Pre-trip/en route	6.586	1.212
4	Time	Adherence to schedule	Punctuality	En route	6.545	1.189
5	Information	Travel information normal conditions	About fare	Pre-trip	6.523	1.194
6	Availability	Operation	Frequency	Pre-trip	6.496	1.192
7	Comfort	Seating and personal space	At B/A points	En route	6.478	1.226
8	Accessibility	Ticketing availability	Validation	En routen/in vehicle	6.433	1.205
9	Information	Travel information normal conditions	About type of ticket	Pre-trip	6.422	1.198
10	Information	General information	About security	In vehicle	6.420	1.205
11	Accessibility	Ticketing availability	Acquisition on network	Pre-trip/in vehicle	6.402	1.227
12	Information	Travel information normal conditions	About route	Pre-trip/en route	6.379	1.216
13	Customer care	Staff	Appearance	In vehicle	6.375	1.208
14	Information	General information	About accessibility	Pre-trip	6.372	1.232
15	Time	Length of trip time	In vehicle	In vehicle	6.369	1.196
16	Customer care	Assistance	*For customers needing help*	Pre-trip/en route	6.364	1.180
17	Customer care	Ticketing options	Payment options	Pre-trip	6.358	1.216
18	Information	General information	About sources of information	Pre-trip/en route	6.352	1.215
19	Comfort	Ambient conditions	Noise	In vehicle	6.315	1.221
20	Comfort	Ambient conditions	Cleanliness	En routen/in vehicle	6.284	1.224
21	Comfort	Seating and personal space	In vehicle	In Vehicle	6.270	1.230
22	Information	General information	About travelling time	Pre-trip/en route	6.263	1.220
23	Availability	Operation	Vehicle load factor	In vehicle	6.262	1.233
24	Comfort	Useability of passenger facilities	*At B/A points*	Pre-trip	6.258	1.221
25	Comfort	Complementary facilities	*Toilets/washing*	Pre-trip	6.231	1.233
26	Security	Freedom from crime	Staff/police presence	Pre-trip/en route	6.227	1.221

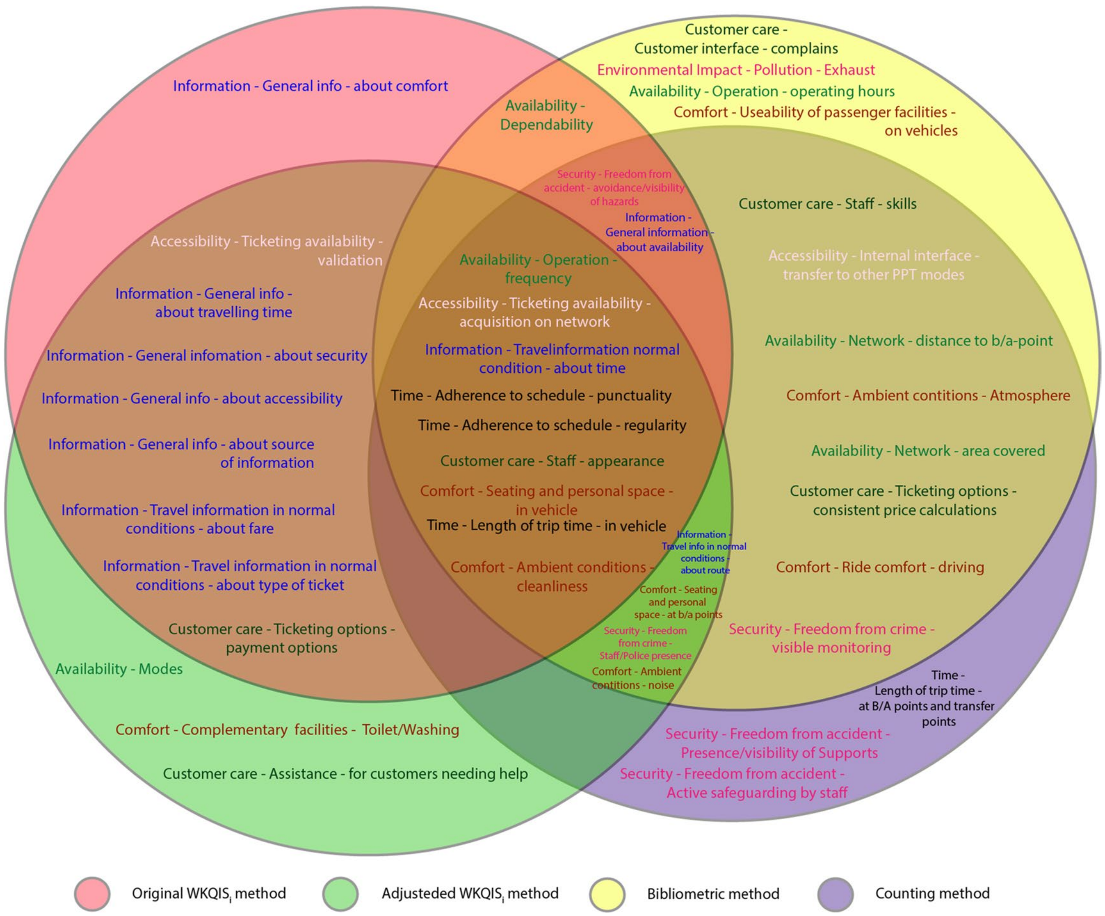


Fig. 4 List of top KQI provided by 4 alternative methods. (Color figure online)

6.2 Link with Previous Research

So far, a set of 26 O_WKQIS_i has been detected in the previous section and compared with 26 A_WKQIS_i. It has been shown that the two criteria share 22 up to 26 most important KQI and that the adjustment process is able to strengthen the scores obtained for each indicator, but it does not overturn their importance. To further investigate about the real importance of the set of main indicators identified by the two scoring methods, in the following, we evaluate if the 26 best ranked KQI are considered equally important in the literature of the quality of transit services. To this purpose, we define a new set of top 26 KQI through a citation score (the counting method) obtained by counting how many times a specific indicator appears in the sources (journal papers, conference papers and technical reports) listed in Table 1. At the same time, a thorough analysis has been carried out with respect to indicators cited in journal papers. For them, a bibliometric ranking method has been used. It considers:

(a) the number of occurrences of an indicator in scientific journals;
(b) the number of journals reporting a specific indicator and ranked in the first (or in the second) quartile of the list obtained from Scimago Institutions Rankings[19] for the category "Transportation";
(c) the number of citable documents published in articles, reviews and conference papers in the three previous years (citable documents—3 years). All the indicators are ranked with respect to the four bibliometric indicators and the final bibliometric rank is the average rank.

The communalities among the four considered methods are represented in Fig. 4, which is self-explanatory. Moreover, for sake of easy, a different color is adopted to edit each service aspect and related indicators according to the classification of CEN/TC 320 (2002).

The bibliometric ranking top 26 KQI set has many communalities with the citation score of the top 26 KQI set (19 up to 26 indicators are common to the two methods) and the overlapping area among the four methods allows us to identify a restricted set of 9 KQI that are common to all methods. This set includes one indicator concerning Accessibility, Availability, Information and Customer Care respectively, three indicators concerning Time and two indicators concerning Comfort.

These results are consistent with Eboli and Mazzulla (2015) which have found that information, cleanliness, and service characteristics like punctuality and frequency of ride have the highest positive effect on service quality. Moreover, these indicators partially differ from dell'Olio et al. (2011) considering that load factor is not ranked within the 9 common KQI. However, a different method of data collection is adopted.

Finally, it is noteworthy that there are four indicators concerning complaints (Customer Facilities), Operating hour (Availability), Usability of passenger facilities on vehicle (Comfort) and exhaust (Environmental Impact) that are relevant from a bibliometric viewpoint but not for the other three methods.

[19] www.scimagojr.com.

7 Conclusion and Research Perspectives

Nowadays, the interest of experts in public transport service studies has been focused on the monitoring of quality. In this domain, the attention of key quality indicators (KQI) depends on their ability to provide informative signals to point out excellences and criticalities in transit services. Therefore, the choice of suitable KQI represents a crucial element to monitoring the quality of services.

Previous research has focused largely on the development of models and methods on how to measure and manage KQI. However, it has seldom investigated coverage-bias adjusted methods for the identification and selection of top KQI. Moreover, to the best of the authors' knowledge, in public transport companies, the monitoring of quality is performed by ad hoc KQI rather than by those derived from a clear and well-organised method.

To account for these points properly, this paper contributes to the growing literature on transit service quality in a twofold manner:

- A novel integrated approach (IA) is proposed to derive top KQI through an algorithmic approach starting from a long list of indicators describing the quality of transit services. IA gathers data on importance of components and attributes as well as marks on indicators from a web-based international survey among experts. Next, IA adjusts outcomes by Monte Carlo simulation methods if coverage-bias occurs and points out the most promising set.
- A comparison between the outcomes of KQI obtained from IA and two different additional sources (i.e., the count of the occurrences of each indicator in scientific publications and the ranking of bibliometric indicators) is provided. A restricted and relevant set of 9 overlapping KQI is identified by merging these outcomes.

This paper extends the original research design whose preliminary results are reported in Barabino et al. (2019) as it considers additional data, both survey and bibliometric, that are used in the statistical analysis.

As for the implications for research and practice deriving from the findings of the present study, the content of this paper is of interest to academics, practitioners, and subsidising agencies. Indeed, this paper sheds new light on a research area that has been largely neglected. Moreover, it is of interest for practitioners of transit industry needing to improve the service quality on routes for benchmarking and/or quality certification purposes according to the European norms (i.e., CEN/TC 320 2002, 2006). In addition, there is a strong interest of subsidising agencies in order to revise the management contracts with PTCs. For instance, they could re-design the service contract in order to include a set of KQI to be monitored in the recent quality incentive-based contracts (e.g., Barry 2005; Hensher and Houghton 2004; Hensher and Stanley 2003; STIB Brussels 2011).

Nevertheless, the present research presents two important limitations. First, since its aim was finding a restricted set of indicators to be used by any organisation worldwide, the research design has not considered the possible trade-off between the specificity of the local conditions and the generality or comparability of the selected indicators. In fact, although the definitions of the selected indicators are widely accepted at high level, local conditions may affect the routes for which these indicators are measured and thus reflect the specificity of the context. Second, although the response rate of the survey is consistent with that of other studies using web-based surveys, we are aware that the achievement of a larger response rate would have further enforce the findings of our research. To account for both limitations, in future research we plan to survey a large sample of experts from not more than one or two countries to effectively verify if a restricted set of important KQI can be identified as well as how local conditions might reflect the choice of the indicators.

To conclude, two further future perspectives are suggested. First, this method can be simply transferred to other transportation modes (e.g., air, maritime, etc.) as well as other general services. Second, a more challenging research based on social web and advancements on web semantic could involve users instead of experts to build the list of KQI. This is possible owing to the feedbacks obtained by users on social media of the transportation companies (e.g., Facebook, Twitter, Instagram). In this way, new KQI could be derived or the old ones confirmed.

Acknowledgements This work was partially supported by the Italian Ministry of University and Research (MIUR), within the Smart City framework (project: PON04a2_00381 “CAGLIARI2020”) and by Regione Autonoma della Sardegna (IT) (Grant Name: Programmazione Unitaria 2007/2013—P.O. FESR 2007/2013, Interventi a sostegno della competitività e dell’innovazione 2016-2018).

Appendix

See Tables 9, 10, 11 and 12.

Table 9 The long list of KQI. *Source*: CEN TC 320 2002

Service aspect	Indicators	Sub-indicator
1. Availability It includes indicators in terms of geography, time, frequency and transport mode	1.1. Modes	
	1.2. Network	1.2.1. Distance to b/a points
		1.2.2. Need for transfer
		1.2.3. Area covered
	1.3. Operation	1.3.1. Operating hours
		1.3.2. Frequency
		1.3.3. Vehicle load factor
	1.4. Suitability	
	1.5. Dependability	
2. Accessibility It includes indicators related to access to the public transport system as well as interface with other transport mode	2.1. External interface	2.1.1. To pedestrians
		2.1.2. To cyclists
		2.1.3. To taxi users
		2.1.4. To private car users
	2.2. Internal interface	2.2.1. Entrance/exits
		2.2.2. Internal movement
		2.2.3. Transfer to other PPT modes
	2.3. Ticketing availability	2.3.1. Acquisition on network
		2.3.2. Acquisition off network
		2.3.3. Validation
3. Information It includes indicators related to the systematic provision of knowledge about the public transport system to assist the customer in the planning and execution of trips	3.1. General information	3.1.1. About availability
		3.1.2. About accessibility
		3.1.3. About sources of information

Table 9 (continued)

Service aspect	Indicators	Sub-indicator
		3.1.4. About travelling time
		3.1.5. About customer care
		3.1.6. About comfort
		3.1.7. About security
		3.1.8. About environmental impact
	3.2. Travel information normal conditions	3.2.1. Street directions
		3.2.2. b/a point identification
		3.2.3. Vehicle direction signs
		3.2.4. About route
		3.2.5. About time
		3.2.6. About fare
		3.2.7. About type of ticket
	3.3. Travel information abnormal conditions	3.3.1. About current/forecast network status
		3.3.2. About alternative available
		3.3.3. About refund/redress
		3.3.4. About suggestions & complaints
		3.3.5. About lost property
4. Time It includes indicators relevant to the planning and execution of trips	4.1. Length of trip time	4.1.1. Trip planning
		4.1.2. Access/egress
		4.1.3. At b/a points at transfer points
		4.1.4. In vehicle
	4.2. Adherence to schedule	4.2.1. Punctuality
		4.2.2. Regularity

Table 9 (continued)

Service aspect	Indicators	Sub-indicator
5. Customer care It includes indicators introduced to influence the closest practicable match between the standard service and the requirements of any individual customer	5.1. Commitment	5.1.1. Customer orientation
		5.1.2. Innovation and initiative
	5.2. Customer interface	5.2.1. Enquiries
		5.2.2. Complaints
		5.2.3. Redress
	5.3. Staff	5.3.1. Availability
		5.3.2. Commercial attitude
		5.3.3. Skills
		5.3.4. Appearance
	5.4. Assistance	5.4.1. At service interruptions
		5.4.2. For customers needing help
	5.5. Ticketing options	5.5.1. Flexibility
		5.5.2. Concessionary tariffs
		5.5.3. Through ticketing
		5.5.4. Payment options
		5.5.5. Consistent price calculations
6. Comfort It includes indicators introduced for making the trip relaxing and pleasurable	6.1. Useability of passenger facilities	6.1.1. At b/a points
		6.1.2. On vehicles
	6.2. Seating and personal space	6.2.1. In vehicle
		6.2.2. At b/a points
	6.3.Ride comfort	6.3.1. Driving
		6.3.2. Starting/stopping
		6.3.3. External factors
	6.4. Ambient conditions	6.4.1. Atmosphere
		6.4.2. Weather protection

Table 9 (continued)

Service aspect	Indicators	Sub-indicator
		6.4.3. Cleanliness
		6.4.4. Brightness
		6.4.5. Congestion
		6.4.6. Noise
		6.4.7 Other undesired activity
	6.5. Complementary facilities	6.5.1. Toilets/washing
		6.5.2. Luggage and other objects
		6.5.3. Communication
		6.5.4. Refreshments
		6.5.5. Commercial services
		6.5.6. Entertainment
	6.6. Ergonomy	6.6.1. Ease of movement
		6.6.2. Furniture design
7. Security It includes indicators introduced for providing a sense of personal protection experienced by customers	7.1. Freedom from crime	7.1.1. Preventivate design
		7.1.2. Lighting
		7.1.3. Visible monitoring
		7.1.4. Staff/police presence
		7.1.5. Identified help points
	7.2. Freedom from accident	7.2.1. Presence/visibility of supports, e.g. handrails
		7.2.2. Avoidance/visibility of hazards
		7.2.3. Active safeguarding by staff
	7.3. Emergency management	7.3.1. Facilities and plans

Table 9 (continued)

Service aspect	Indicators	Sub-indicator
8. Environmental impact It includes indicators related to the environment where the transit service is operated	8.1. Pollution	8.1.1. Exhaust
		8.1.2. Noise
		8.1.3. Visual pollution
		8.1.4. Vibration
		8.1.5. Dust and dirt
		8.1.6. Odour
		8.1.7. Waste
		8.1.8. Electromagnetic interference
	8.2. Natural resources	8.2.1. Energy
		8.2.2. Space
	8.3. Infrastructure	8.3.1. Effect of vibration
		8.3.2. Wear on road/rail etc.
		8.3.3. Demands on available resources
		8.3.4. Disruption by other activities

The long list of service aspects, indicators and sub-indicators, according to European Norm EN 13816:2002 (CEN TC 320 2002). The numerical code is the original one provided by this norm

Table 10 Distribution of observed weights of methodological and relevance to quality components

Component	Symbol	Mean weight	Standard deviation	Coefficient of variation (%)
Methodological features	$\bar{M}$	0.368	0.269	73.1
Academics		0.355	0.262	73.9
Asia–Pacific		0.367	0.298	81.3
Australia and New Zealand		0.375	0.292	77.8
Europe		0.400	0.274	68.4
North and South America		0.315	0.259	82.2
Practitioners		0.444	0.276	62.2
Asia–Pacific		0.650	0.212	32.6
Australia and New Zealand		0.569	0.452	79.3
Europe		0.437	0.267	61.2
North and South America		0.163	0.053	32.6
Relevance to transit service quality	$\bar{Q}$	0.632	0.269	42.6
Academics		0.645	0.262	40.7
Asia–Pacific		0.633	0.298	47.1
Australia and New Zealand		0.625	0.292	46.7
Europe		0.600	0.274	45.6
North and South America		0.685	0.259	37.8
Practitioners		0.556	0.276	49.6
Asia–Pacific		0.350	0.212	60.6
Australia and New Zealand		0.431	0.452	104.9
Europe		0.563	0.267	47.5
North and South America		0.838	0.053	6.3

Table 11 Distribution of observed weights of methodological components

Attribute	Symbol	Mean weight	Standard deviation	Coefficient of variation (%)
Measurability	$\overline{w_1}$	0.261	0.177	67.9
Academics		0.299	0.197	65.7
Asia–Pacific		0.166	0.103	62.0
Australia and New Zealand		0.190	0.167	87.8
Europe		0.411	0.228	55.5
North and South America		0.324	0.179	55.4
Practitioners		0.231	0.140	60.4
Asia–Pacific		0.194	0.061	31.2
Australia and New Zealand		0.279	0.180	64.5
Europe		0.226	0.154	68.2
North and South America		0.261	0.083	31.6
Ease of availability	$\overline{w_2}$	0.285	0.156	54.6
Academics		0.256	0.158	61.8
Asia–Pacific		0.418	0.176	42.1
Australia and New Zealand		0.335	0.261	78.0
Europe		0.181	0.077	42.7
North and South America		0.209	0.078	37.5
Practitioners		0.286	0.155	54.1
Asia–Pacific		0.270	0.180	66.6
Australia and New Zealand		0.253	0.003	1.4
Europe		0.279	0.168	60.2
North and South America		0.394	0.147	37.2
Speed of availability	$\overline{w_3}$	0.146	0.120	82.2
Academics		0.137	0.088	64.6
Asia–Pacific		0.171	0.091	53.1
Australia and New Zealand		0.201	0.106	53.0
Europe		0.152	0.095	62.4
North and South America		0.089	0.053	59.5
Practitioners		0.175	0.154	88.0
Asia–Pacific		0.148	0.110	73.9
Australia and New Zealand		0.047	0.010	21.2
Europe		0.179	0.166	92.5
North and South America		0.293	0.102	34.9
Interpretability	$\overline{w_4}$	0.308	0.175	56.7
Academics		0.309	0.180	58.4
Asia–Pacific		0.244	0.172	70.2
Australia and New Zealand		0.274	0.182	66.4
Europe		0.256	0.176	68.6
North and South America		0.379	0.181	47.9
Practitioners		0.308	0.170	55.3
Asia–Pacific		0.388	0.229	59.0
Australia and New Zealand		0.421	0.167	39.6
Europe		0.316	0.154	48.8
North and South America		0.051	0.038	74.5

Table 12 Distribution of adjusted weights of quality components

Attribute	Symbol	Mean weight	Standard deviation	Coefficient of variation (%)
Integration users-companies	$\overline{w_5}$	0.197	0.098	49.8
Academics		0.180	0.089	49.3
Asia–Pacific		0.158	0.092	58.1
Australia and New Zealand		0.235	0.087	37.0
Europe		0.133	0.086	64.9
North and South America		0.196	0.082	42.1
Practitioners		0.190	0.112	59.0
Asia–Pacific		0.141	0.015	10.4
Australia and New Zealand		0.085	0.028	32.4
Europe		0.214	0.122	57.1
North and South America		0.152	0.009	5.6
User orientation	$\overline{w_6}$	0.360	0.207	57.6
Academics		0.397	0.203	51.2
Asia–Pacific		0.461	0.187	40.6
Australia and New Zealand		0.451	0.132	29.3
Europe		0.494	0.228	46.1
North and South America		0.291	0.183	63.1
Practitioners		0.299	0.204	68.2
Asia–Pacific		0.329	0.096	29.2
Australia and New Zealand		0.125	0.094	75.6
Europe		0.346	0.209	60.4
North and South America		0.072	0.037	50.6
Subjective and objective measurability	$\overline{w_7}$	0.248	0.170	68.4
Academics		0.256	0.180	70.4
Asia–Pacific		0.208	0.188	90.1
Australia and New Zealand		0.193	0.128	66.3
Europe		0.240	0.186	77.5
North and South America		0.309	0.195	63.1
Practitioners		0.247	0.157	63.9
Asia–Pacific		0.374	0.209	55.9
Australia and New Zealand		0.403	0.285	70.6
Europe		0.215	0.139	64.7
North and South America		0.219	0.085	38.8
Amount of passengers	$\overline{w_8}$	0.195	0.155	79.5
Academics		0.167	0.111	66.3
Asia–Pacific		0.173	0.105	60.9
Australia and New Zealand		0.121	0.095	78.3
Europe		0.132	0.092	69.4
North and South America		0.205	0.126	61.6
Practitioners		0.264	0.191	72.2
Asia–Pacific		0.156	0.098	62.8
Australia and New Zealand		0.387	0.218	56.4
Europe		0.226	0.172	76.3
North and South America		0.557	0.113	20.3

References

AFNOR Certification. NF 281. (2005a). Services de transport urbain de voyageurs. Règlement de certification générique. Saint-Denis la Plaine Cedex, France. Technical report.

AFNOR Certification. NF 286. (2005b). Services de transport urbain de voyageurs. Règlement de certification spécifiques. Saint-Denis la Plaine Cedex, France. Technical report.

Ayadi, A., & Hammami, S. (2015). Analysis of the technological features of regional public transport companies: The Tunisian case. *Public Transport, 7*(3), 429–455.

Aydin, N., Celik, E., & Gumus, A. T. (2015). A hierarchical customer satisfaction framework for evaluating rail transit systems of Istanbul. *Transportation Research Part A, 77,* 61–81.

Badami, M. G., & Haider, M. (2007). An analysis of public bus transit performance in Indian cities. *Transportation Research Part A, 41,* 961–981.

Balcome, R., Mckett, R., Pulley, N., Preston, J., Shires, J., Titheridge, H., et al. (2004). The demand for public transport: A practical guide. TRL report, TRL 593, TRL Limited.

Barabino, B. (2018). Automatic recognition of "low-quality" vehicles and bus stops in bus services. *Public Transport, 10*(2), 257–289.

Barabino, B., Conversano, C., Cabras, N. A., & Fantola, M. (2019). Selecting key quality indicators in public transport systems using a robust method. In F. A. Ortega (Ed.), *WIT transactions on the built environment* (Vol. 182, pp. 73–84). Southampton: WIT Press.

Barabino, B., & Deiana, E. (2013). On the attributes and influencing factors of end-users quality perceptions in urban transport: An exploratory analysis. *Procedia: Social and Behavioral Sciences, 87,* 18–30.

Barabino, B., Deiana, E., & Mozzoni, S. (2013). The quality of public transport service: The 13816 standard and a methodological approach to an Italian case. *Ingegneria Ferroviaria, 68*(5), 475–499.

Barabino, B., Deiana, E., & Mozzoni, S. (2015a). A framework to measure transit service quality areas to be managed. *International Journal of Productivity and Quality Management, 16*(4), 390–415.

Barabino, B., Deiana, E., & Tilocca, P. (2011). Urban transport management and customer perceived quality: A case study in the metropolitan area of Cagliari, Italy. *Theoretical and Empirical Researches in Urban Management, 6*(1), 19.

Barabino, B., Deiana, E., & Tilocca, P. (2012). Measuring service quality in urban bus transport: A modified SERVQUAL approach. *International Journal of Quality and Service Sciences, 4*(3), 238–252.

Barabino, B., & Di Francesco, M. (2016). Characterizing, measuring, and managing transit service quality. *Journal of Advanced Transportation, 50*(5), 818–840.

Barabino, B., Di Francesco, M., & Mozzoni, S. (2015b). Rethinking bus punctuality by integrating automatic vehicle location data and passenger patterns. *Transportation Research Part A, 75,* 84–95.

Barabino, B., Lai, C., Casari, C., Demontis, R., & Mozzoni, S. (2017). Rethinking transit time reliability by integrating automated vehicle location data, passenger patterns, and web tools. *IEEE Transactions on Intelligent Transportation Systems, 18*(4), 756–766.

Barry, J. (2005). Transport for London. Prepared for master in "European management training in transport and sustainable mobility". Dublin.

Beirão, G., & Cabral, J. S. (2007). Understanding attitudes towards public transport and private car: A qualitative study. *Transport Policy, 14*(6), 478–489.

Berry, L. L., Zeithaml, V. A., & Parasuraman, A. (1990). Five imperatives for improving service quality. *Sloan Management Review Summer, 31*(4), 29–38.

Cantwell, M., Caulfield, B., & O'Mahony, M. (2009). Examining the factors that impact public transport commuting satisfaction. *Journal of Public Transportation, 12*(2), 1–21.

Castillo, H., & Pitfield, D. (2010). ELASTIC: A methodological framework for identifying and selecting sustainable transport indicators. *Transportation Research Part D, 15*(4), 179–188.

CEN/TC 320. (2002). Transportation—Logistics and services. European Standard EN 13816: Public passenger transport—Service quality definition, targeting and measurement. European Committee for Standardization, Brussels. Technical report.

CEN/TC 320. (2006). Transportation—Logistics and services, European Standard EN 15140: Public passenger transport—Basic requirements and recommendation for systems that measure delivered service quality, European Committee for Standardization, Brussels. Technical report.

Chen, X., Yu, L., Zhang, Y., & Gou, J. (2009). Analysing urban bus service reliability at the stop, route, and network levels. *Transportation Research Part A, 43*(8), 722–734.

Dawes, J. (2008). Do data characteristics change according to the number of scale points used? *International Journal of Market Research, 50*(1), 61–77.

de Oña, J., & de Oña, R. (2014). Quality of service in public transport based on customer satisfaction surveys: A review and assessment of methodological approaches. *Transportation Science, 49*(3), 605–622.

de Oña, J., de Oña, R., Eboli, L., & Mazzulla, G. (2016). Index numbers for monitoring transit service quality. *Transportation Research Part A: Policy and Practice, 84,* 18–30.
de Steiguer, J. E., Duberstein, J., & Lopes, V. (2003). The analytic hierarchy process as a means for integrated watershed management. In K. G. Renard (Ed.), *First interagency conference on research on the watersheds* (pp. 736–740). Benson: U.S. Department of Agriculture, Agricultural Research Service.
dell'Olio, L., Ibeas, A., & Cecin, P. (2011). The quality of service desired by public transport users. *Transport Policy, 18*(1), 217–227.
Dever, J. A., Rafferty, A., & Valliant, R. (2008). Internet surveys: Can statistical adjustments eliminate coverage bias? *Survey Research Methods, 2*(2), 47–62.
Dhakal, S., & Imura, H. (2003). Policy-based indicator systems: Emerging debates and lessons. *Local Environment, 8*(1), 113–119.
Doran, G. T. (1981). There's a SMART way to write managements' goals and objectives. *Management Review, 70,* 35–36.
Dyer, J. S. (1990). Remarks on the analytic hierarchy process. *Management Science, 36*(3), 249–258.
Eboli, L., Forciniti, C., & Mazzulla, G. (2018). Spatial variation of the perceived transit service quality at rail stations. *Transportation Research Part A, Policy and Practice, 114,* 67–83.
Eboli, L., & Mazzulla, G. (2007). Service quality attributes affecting customer satisfaction for bus transit. *Journal of Public Transportation, 10*(3), 21–34.
Eboli, L., & Mazzulla, G. (2009). A new customer satisfaction index for evaluating transit service quality. *Journal of Public Transportation, 12*(3), 21–37.
Eboli, L., & Mazzulla, G. (2010). How to capture the passengers point of view on a transit service through rating and choice options. *Transport Reviews, 30*(4), 435–450.
Eboli, L., & Mazzulla, G. (2011). A methodology for evaluating transit service quality based on subjective and objective measures from the passenger's point of view. *Transport Policy, 18,* 172–181.
Eboli, L., & Mazzulla, G. (2012). Performance indicators for an objective measure of public transport service quality. *European Transport, 51,* 1–21.
Eboli, L., & Mazzulla, G. (2015). Relationships between rail passengers' satisfaction and service quality: A framework for identifying key service factors. *Public Transport, 7*(2), 185–201.
Forman, E. H., & Gass, S. I. (2001). The analytical hierarchy process-an exposition. *Operations Research, 49*(4), 469–487.
Friman, M., Advardsson, B., & Gärling, T. (2001). Frequency of negative critical incidents and satisfaction with public transport services. *Journal of Retailing and Consumer Services, 8,* 95–104.
Furth, P. G., Hemily, B., Muller, T. H. J., & Strathman, J. (2006). Using archived AVL-APC data to improve transit performance and management. TRB, Washington, DC, technical report TRCP 113.
Guirao, B., García-Pastor, A., & López-Lambas, M. E. (2016). The importance of service quality attributes in public transportation: Narrowing the gap between scientific research and practitioners' needs. *Transport Policy, 49,* 68–77.
Harker, P. T., & Vargas, L. G. (1990). Reply to "remarks on the analytic hierarchy process" by JS Dyer. *Management Science, 36*(3), 269–273.
Hassan, M. N., Hawas, J. E., & Ahmed, K. (2013). A multi-dimensional framework for evaluating the transit service performance. *Transportation Research Part A, 50,* 47–61.
Heilig, L., & Voß, S. (2015). A scientometric analysis of public transport research. *Journal of Public Transportation, 18*(2), 111–141.
Hensher, D. A. (2007). *Bus transport: Economics, policy, and planning*. New York: JAI Press.
Hensher, D. A. (2014). The relationship between bus contract costs, user perceived service quality and performance assessment. *International Journal of Sustainable Transportation, 8*(1), 5–27.
Hensher, D. A. (2015). Using contracted assets to undertake non-contracted services as a way to improve cost efficiency under negotiated or tendered bus contracts. *Journal of Strategic Contracting and Negotiation, 1*(2), 111–128.
Hensher, D. A., & Houghton, J. (2004). Performance-based quality contracts for the bus sector: delivering social and commercial value for money. *Transportation Research Part B, 38,* 123–146.
Hensher, D. A., & Stanley, J. (2003). Performance-based quality contracts in bus service provision. *Transportation Research Part A, 37,* 519–538.
Hirschhorn, F., Veeneman, W., & van de Velde, D. (2018). Inventory and rating of performance indicators and organisational features in metropolitan public transport: A worldwide Delphi survey. *Research in Transportation Economics, 69,* 144–156.
Kittelson & Associates. (2003). *Transit cooperative research program & transit development corporation. A guidebook for developing a transit performance-measurement system* (Vol. 88). Washington: Transportation Research Board.

Kittelson & Associates Inc, KFH Group Inc, Parsons Brinckerhoff Quade & Douglass Inc, & Zaworski, K. H. (2003). *Transit capacity and quality of service manual* (2nd ed.). Washington, DC: TRB.
Liekendael, J. C., Furth, P. G., & Muller, T. H. J. (2006). Service quality certification in Brussels, Belgium. A quality process with teeth. *Transportation Research Record, 1955,* 88–95.
Lin, J., Wang, M. L., & Barnum, P. D. (2008). A quality control framework for bus schedule reliability. *Transportation Research Part E, 44*(6), 1086–1098.
Macharis, C., Springael, J., De Brucker, K., & Verbeke, A. (2004). PROMETHEE and AHP: The design of operational synergies in multicriteria analysis: Strengthening PROMETHEE with ideas of AHP. *European Journal of Operational Research, 153*(2), 307–317.
Mahmoud, M. M., Hine, J., & Kashyap, A. (2011). Bus transit service quality monitoring in UK: A methodological framework. In *Proceedings of the ITRN 2011*. www.itrn.ie/uploads/sesD2_ID75.pdf. Retrieved September 28, 2016.
Millet, I., & Wedley, W. C. (2002). Modelling risk and uncertainty with the analytic hierarchy process. *Journal of Multi-Criteria Decision Analysis, 11,* 97–107.
Moffat, A. (2014). The evolution of bus contracts in London. Transport for London. http://embarqbrasil.org/sites/default/files/9.%20Alex%20Moffat%20-%20London.pdf. Retrieved September 10, 2016.
MORPACE International Inc, Cambridge Systematics Inc. (1999). A handbook for measuring customer satisfaction and service quality. TRB, TCRP report 47, National Research Council, Washington DC.
Nathanail, E. (2008). Measuring the quality of service for passengers on the Hellenic railways. *Transportation Research Part A, 42,* 48–66.
Oguztimur, S. (2011). Why fuzzy analytic hierarchy process approach for transport problems? 51st congress of the European Regional Science Association, 30 August–3 September 2011, Barcelona, Spain.
Parasuraman, A., Berry, L. L., & Zeithaml, V. A. (1991). Refinement and reassessment of the SERVQUAL scale. *Journal of Retailing, 67*(4), 420–450.
Prioni, P., & Hensher, D. A. (2000). Measuring service quality in scheduled bus services. *Journal of Public Transportation, 3*(2), 51–74.
Ramanathan, R. (2001). A note on the use of the analytic hierarchy process for environmental impact assessment. *Journal of Environmental Management, 63,* 27–35.
Saaty, T. L. (1980). *The analytic hierarchy process*. New York: McGraw-Hill.
Saaty, R. W. (1987). The analytic hierarchy process—What it is and how it is used. *Mathematical Modelling, 9*(3–5), 161–176.
Saaty, T. L. (1990). An exposition of the AHP in reply to the paper "remarks on the analytic hierarchy process". *Management Science, 36*(3), 259–268.
Sheth, C., Triantis, K., & Teodorovic, D. (2007). Performance evaluation of bus routes: A provider and passenger perspective. *Transportation Research Part E, 43,* 453–478.
Sivo, S. A., Saunders, C., Chang, Q., & Jiang, J. J. (2006). How low should you go? Low response rates and the validity of inference in IS questionnaire research. *Journal of the Association for Information Systems, 7*(6), 351–414.
STIB Bruxelles. (2011). Manuel de maitrise de la qualitè de service. Technical report.
Stradling, S., Anable, J., & Carreno, M. (2007). Performance, importance and user disgruntlement: A six method for measuring satisfaction with travel modes. *Transportation Research Part A, 41*(1), 98–106.
Too, L., & Earl, G. (2010). Public transport service quality and sustainable development: A community stakeholder perspective. *Sustainable Development, 18*(1), 51–61.
Tyrinopoulos, Y., & Antoniou, C. (2008a). A complete methodology for the quality control of passenger services in the public transport business. *European Transport, 15,* 1–16.
Tyrinopoulos, Y., & Antoniou, C. (2008b). Public transit user satisfaction: Variability and policy implications. *Transport Policy, 15,* 260–272.
Wang, T. C., & Lee, H. D. (2009). Developing a fuzzy TOPSIS approach based on subjective weights and objective weights. *Experts Systems with Applications, 36*(5), 8980–8985.
Zelany, M. (1974). A concept of compromise solutions and the method of the displaced ideal. *Computers and Operations Research, 1*(3–4), 479–496.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.