Transport quality profiles of European cities based on a multidimensional set of satisfaction ratings indicators

# Transport Quality Profiles of European Cities Based on a Multidimensional Set of Satisfaction Ratings Indicators

Marco Diana, André Duarte, Miriam Pirra

28th February 2017



***Preferred citation***: this document may be cited directly referring to the above mentioned final published version:

**Diana, M., Duarte, A. and Pirra, M. (2017) Transport Quality Profiles of European Cities Based on a Multidimensional Set of Satisfaction Ratings Indicators, *Transportation Research Record: Journal of the Transportation Research Board, No. 2643*, pp. 84-92.**

# Transport Quality Profiles of European Cities Based on a Multidimensional Set of Satisfaction Ratings Indicators

**Marco Diana**
DIATI - Department of Environment, Land and Infrastructure Engineering
Politecnico di Torino
Corso Duca degli Abruzzi, 24, 10129 Torino, ITALY
Tel: +39 011 090 5638 Fax: +39 011 090 7699; Email: marco.diana@polito.it

**André Duarte**
VTM Consultores
Av. 25 de Abril de 1974, 23 2ºA, 2795-197 Linda-a-Velha, PORTUGAL
Tel: +351 21 415 96 00 Fax: +351 21 415 96 08; Email: andre.duarte@vtm.pt

**Miriam Pirra, Corresponding Author**
DIATI - Department of Environment, Land and Infrastructure Engineering
Politecnico di Torino
Corso Duca degli Abruzzi, 24, 10129 Torino, ITALY
Tel: +39 011 090 5639 Fax: +39 011 090 7699; Email: miriam.pirra@polito.it

Word count: 6,190 words text + 5 tables/figures x 250 words (each) = 6,190 + 1,250 = 7,440 words

TRR Paper number: 17 - 01894

Submission date: 28th February 2017

**ABSTRACT**

A monitoring system of the perceived quality of transport services at the city level is proposed, based on a set of 92 indicators covering different travel modes and considering the viewpoint of different groups. Special emphasis is given to public transport, bicycles and pedestrians and to travelers with special needs (e.g. commuters, visitors, mobility challenged or communication impaired). Indicators are found through Principal Component Analysis from a set of satisfaction ratings elicited through a survey in eight European cities, implemented within the METPEX research project. Benchmarking values are computed for all indicators to provide the analyst with some initial guidance. All indicators are then visualized through a dashboard that can give policy makers a synthetic overview of the main areas where perceived quality and accessibility are above or below average. At the same time, jointly considering indicators from different dimensions of the evaluation exercise (by travel means, by traveler profile, by phase of the journey experience) provides additional insights on specific issues that would be overlooked in a coarser quality assessment activity.


*Keywords*: Transport indicators, Quality monitoring, Benchmark, Accessibility, Principal Component Analysis

## INTRODUCTION

Quality monitoring of transport services is a burgeoning research field, with an ever-increasing number of works available in the open literature. Yet, despite the massive amount of papers dealing with such topics, the research mainstream in the field is most of the time focusing on some issues or privileging specific viewpoints. In a recent review dealing with public transport quality studies, de Oña and de Oña *(1, p. 616)* note that stakeholders and practitioners often rely on the feedback of descriptive statistics from customer satisfaction surveys in which ratings are given on different service features, while researchers tend to propose models to integrate a larger variety of data into a unique quality index. The former approach can provide an immediate view on the aspects of the service that need improvement, while the latter might ease the comparability across different services or over time.

However, there is not a consensus on the theoretical framework needed to develop a measurement model, concerning for example the definition of quality, the attributes needed to evaluate it, or the opportunity of exclusively considering subjective assessments versus integrating them with objective performance measures. The U.S. Transit Capacity and Quality of Service Manual considers quality as the reflection of “the passenger's perception of transit performance” *(2, p.4-1)*, whereas a transit performance measurement system can include hundreds of indicators to adequately cover all evaluation aspects and the disparate viewpoints of different stakeholders *(3)*. Conversely, the European Standard EN-13816 envisions a “quality loop” where both the provider and user viewpoints are considered *(4)*; however, only some guidelines to develop quality indicators are provided.

Monitoring the quality of transport systems can in principle have two complementary objectives. On a service operator viewpoint, the goal is to increase the number of customers, i.e. of travelers on a given service, and maximize profitability like for any enterprise in a market economy. However, transport authorities and policy makers should have a broader view, jointly assessing the quality of all services in a city (possibly not limited to public transport) and appropriately considering different users groups. Their goal is in fact not the profit maximization but the achievement of some general welfare objectives that justify public interventions (taxes and subsidies) on different transport services. A large part of studies in the literature, such many of as those reviewed by de Oña and de Oña *(1)*, derive their methods from the marketing literature in the private sector, so that the latter viewpoint is often overlooked.

The objective of the present paper is to contribute in filling this research gap by proposing a multimodal quality monitoring system for transport services at the city level, in order to provide policy makers with an effective decision support tool that is based on the feedback from different social groups. This research effort was part of METPEX, a European project aimed at building a tool for the measurement of the quality of the whole traveler experience (www.metpex.eu). Wishing to facilitate the use of the tool among stakeholders, an array of indicators has been built only on the basis of subjective ratings from customer satisfaction surveys, a familiar tool in professional practice. Given the above mentioned lack of agreement on the theory behind quality monitoring even from the much narrower perspective of public transport operators and agencies, no modeling attempt was carried out in the project. An exploratory and data-driven analysis was then implemented, as opposed to confirmatory methodologies that are more commonplace in the scientific literature, where expert judgments firstly define the indicators to be later tested through empirical fieldwork.

It is believed that this approach can give contributions to the advancement of the state of the art under several viewpoints. First and foremost, papers dealing with quality measurement and monitoring of transport services focus only on public transit *(1)*. However, other travel means such as bicycles and feet should be considered in a quality assessment exercise that is relevant for policy and strategic planning. Under certain conditions mainly dealing with land use patterns, active modes can in fact be seen as an additional alternative beyond public transport to substitute motorized individual trips. There is thus an interest for policy makers to understand satisfaction levels also related to active travel means. To the best of the authors' knowledge, no papers in the open literature are specifically dealing with the development of indicators to measure the perceived journey quality of bikers and walkers. Quality indicators for several different travel modes are proposed in METPEX.

One of the central ideas of METPEX is to consider the whole journey experience, as opposed to the travel activity. The journey experience starts with the contemplation stage of making a trip, continues through a series of pre-trip activities such as checking schedules or buying tickets in advance, encompasses the trip itself and extends to any post trip activity, such as filing a complaint. Such broader unit of analysis can better accommodate the intricacies of quality assessment of multimodal trips, and it is also more relevant to stakeholders wishing to link satisfaction levels for different components of the system with the relevant actor. For example, in some institutional settings, information provision is a task carried out by an agency, while the service itself is contracted to a private enterprise and the ground infrastructure is maintained by a publicly-owned enterprise. Some recent studies have considered at least partly this larger perspective (e.g. *5-8*), however restricting themselves to public transport trips. Conversely, the relationship between overall satisfaction for multimodal trips and satisfaction with the single trip stages and modes being used is very complex and deserves further research *(9)*. Some of the below quality indicators are dealing with specific phases of the journey experience, such as walking between transport infrastructures in multimodal trips (bus stops, parking).

It is also well known that satisfaction ratings are heavily depending on subjective conditions, beyond the actual performances of the different services. To consider the viewpoints of different users, one common approach is to perform a market segmentation and to check how quality assessments change according to the specific group *(10-11)*. In the literature, some models consider the heterogeneity of users in defining synthetic quality indices *(12)* or study the effect of some sociodemographic variables on perceived quality *(13-14)*. However, basing the customer profiling only on sociodemographic characteristics is somewhat limiting, since the same individual can have different views on the same trip according to the specific context in which the trip is made. When defining traveler types, METPEX therefore considers situational variables as well, such as the fact of being a visitor or a commuter for a particular trip, to better capture such variability factor. Additionally, the developed system of quantitative indicators tries to embed the viewpoint of physically challenged and disadvantaged travelers.

The paper unfolds as follows. Experimental activities aimed at gathering satisfaction ratings in eight European cities are described in the following section, while the derivation of indicators from such data is illustrated in the "Methodology to define and compute the indicators" section. Once defined the indicators, their values are computed in each of the eight cities to create some benchmarks that might useful as reference values. The final outcome of the research is a visualization tool jointly displaying the values taken by several different indicators. This tool can help stakeholders to quickly identify what are the views of different kinds of travelers and what are the areas of improvement within the overall transport system. An example on the use of the tool in one of the eight cities that tested it during the project is then illustrated.

## EMPIRICAL SETTINGS TO GATHER SATISFACTION RATING

The fulfilment of the above mentioned METPEX project objectives involved the organization of a travel satisfaction survey in eight European cities: Bucharest (Romania), Coventry (UK), Dublin (Ireland), Grevena (Greece), Rome (Italy), Stockholm (Sweden), Valencia (Spain) and Vilnius (Lithuania). The questionnaire structure followed the above mentioned idea of building indicators by jointly considering several possible evaluation frameworks and following a data-driven approach rather than some expert judgment. Therefore, the first phase of the survey design implied the formulation of 417 satisfaction rating questions, which tried to cover in details all different aspects that might influence the overall perceived quality of the single journey experience that was investigated in the survey. Such questions came out from a brainstorming exercise within the project team, made up of people with different backgrounds (social scientists, transport engineers and planners, IT programmers, professors, local authorities etc.).

An experimental design was then carried out to select about 60 items from this longer list to be presented to each respondent, plus some baseline questions related to the sociodemographic characteristics of the traveler and to the features of the journey under investigation. Out of these 60 satisfaction ratings, 21 were elicited from all respondents to have a common base and they were named “Tier-1”. An additional mode-specific set of about 12 questions changed according to the travel mode that was used for the longest time to complete the journey under investigation. To capture the disparate viewpoints of different kinds of travelers, each respondent was assigned to one of the following 11 categories according to the answers given in the sociodemographic section of the survey: women, commuters, low income, aged over 65, aged under 24, travelling with children, travelling with dependents, communication impaired, mobility restricted, rural dwellers and visitors. A third user-group specific set of about 12-15 ratings was then assigned to each respondent according to such profiles.

It is clear that these categories are not mutually excluding. Therefore, each of the above eight sites could judgmentally pre-assign a weight to all mode-specific and user group-specific categories that was proportional to the estimated frequency of the category in their universe (so that, for example, “woman” and “car” could have larger weights than “rural dweller” and “bike”). During the interview all categories that apply to a given respondent were identified, on the basis of his/her answers to the above mentioned baseline questions. Then, respondents were assigned to one user group category and one mode category with an outcome probability that is the reciprocal of aforementioned weight. For example, a mobility restricted woman would have a much higher probability to answer to questions related to the “mobility restricted” group, which is assumed to be less numerous in the test site.

The fourth and last set of ratings was randomly assigned from a pool of 21 “Tier-2” sets, each set being related to one of the above 21 Tier-1 questions, and it involved the evaluation of very specific aspects of the journey experience (e.g. “Width of aisles/corridors in vehicles” or “Accuracy of the fare information on the web/apps”).

For both mode-specific and traveler profile-specific questions sets, an unequal probability design was implemented to oversample less common groups. Overall, 4,665 valid responses have been collected in the eight cities between September and November 2014, with a site-specific sample size ranging from 319 to 993. Additional observations were gathered by distributing the questionnaire to the network of the International Federation of Motorists, but they are not considered here. The overall sample size was sufficient to separately analyze all the above 11 traveler types except “traveling with dependents”. Detailed descriptive statistics on the data are not reported here to save on space but they are available in Susilo et al. *(15)*.

## METHODOLOGY TO DEFINE AND COMPUTE THE INDICATORS

### Classifications of Variables and Data Reduction Techniques

The very rich dataset that was generated from the previously described survey has been analyzed to discover the underlying patterns of satisfaction ratings for several distinct groups of observations and of variables. The first step was therefore to select both the observations and the satisfaction ratings that were relevant to build different indicators. If, for example, the focus is on bike trips, it is clear that only individuals that used such travel means to complete their trip need to be considered, along with only those satisfaction items that are relevant to such travel mode among all those that were rated by such individuals.

To efficiently select the relevant variables according to the kind of indicator, all the above mentioned 417 satisfaction ratings were classified through the following three different evaluation dimensions: by travel means, by traveler type and by phase of the journey experience *(16)*. Several tags were therefore added to each variable to identify the relevant ambits to which the related question can be referred, irrespective of the point of the questionnaire in which it was located (for example, many Tier-2 questions are only relevant for transit services). However, many of the 417 variables were assessed only through a tiny fraction of the observations for which they are relevant, mainly due to the above introduced survey design. Hence, variables with relatively few observed values were excluded from the following analyses.

The second column of Table 1 lists the tags that were used for each of the above three evaluation dimensions and the third column reports the corresponding number of considered satisfaction ratings variables. Due to the above mentioned exclusion process and to the fact that the variables may have zero, one or several tags within each evaluation dimension, the sum of values in the second column of the table corresponding to “Travel mode”, “User group” or “Journey phase” is different from 417.

Beyond such analysis, an additional set of indicators was constructed by separately considering the answers given to each set of Tier-2 questions among the 21 available. In this case there was no need to tag the variables, since each variable is already linked to one corresponding Tier-1 question. The first column of Table 2 therefore lists the 21 quality components as identified by the Tier-1 questions. Indicators stemming from this analysis are more general, not particularly focusing on specific travel means or traveler kinds, although some of these are in practice only relevant for public transport. It is also worth noting that, through some of the Tier-1 questions, opinions to aspects relevant to travelers with additional needs were asked to the entire sample.

Finally, a specific focus of the project was on the views of mobility disadvantaged people using public transport. The size of the sample was sufficient to run two additional analyses on a subsample only considering such category of travelers, while selecting the variables on the basis of a cross-classification of two of the above evaluation dimensions. Namely, the first analysis was performed on a subset of variables that were tagged with both “mobility restricted” and any of the four transit modes (railway services, underground, tramways and bus services), while the second one rather considered “communication impaired” tags along with any of the four transit modes.

After having identified the different groups of variables through the above three evaluation dimensions, the sets of Tier-2 questions and the cross-classifications, two data analyses were sequentially implemented to find the indicators within each group. Namely, Cronbach alpha was first employed to assess if the observed variance in satisfaction ratings within each group can be explained by a unique latent factor, i.e. to which extent different variables represents the same concept, when compared to others *(17)*. This mostly happened when considering the quality components in Table 2. If this was not the case, then a data reduction technique such Principal Component Analysis (PCA) with varimax rotation was employed to find the minimum set of latent factors that can explain such variance *(18)*. The fourth column of Table 1 and the third column of Tables 2 and 3 report the number of components that were distilled from each group of variables. The detailed definition of such components, i.e. the lists of variables along with their loadings, can be found in *(19)*, along with a discussion on the motivations that led us to prefer PCA over Factor Analysis.

However, not all such components were automatically considered to generate a distinct indicator. Some of them were in fact either rather generic, or the underlying set of variables was not clearly interpretable and were therefore discarded. Concerning indicators specific for travel modes and users group, better results were most of the times found by excluding Tier-2 variables from the analysis. After a considerable testing phase, a final set of 92 quality indicators, as shown in the last column of Tables 1, 2 and 3, was then identified. Non-consecutive numbers of labels (for example, no indicator is named BIKE3 but there is BIKE4) are due to the selection made during such tests *(20)*. Given the focus of the project on public transport and active means, only one indicator for motorized private vehicles has been retained for comparison purposes.

**Stability Analysis**

Data reduction techniques such as Exploratory Factor Analysis or Principal Component Analysis are often criticized, given their purely empirical approach that makes their results hardly generalizable. Since the goal of the project is to set up a replicable method of analysis, some validation activities have been carried out to check if the latent factors represented by the above indicators could be found also in different empirical contexts. While validation of models is routinely carried out in transport research, it is much less commonly encountered in Exploratory Factor Analysis studies, that often rely on samples collected within a unique study area.

In the present study the same survey was carried out in different cities, therefore it was possible to replicate the PCAs that were first done on the whole dataset, only using the observations that are available within one city. The analysis is carried out on those indicators for which a sufficiently large number of observations is available in at least two different cities to proceed with a distinct PCA.

As an example, Table 4 presents the results that can be obtained when the PCA that allowed the identification of indicators RAIL1 and RAIL2 is replicated only on the observations gathered in Rome and Stockholm. The values provided in the table are known as “component loadings”, representing how much each variable “contributes” to an indicator. The comparison of this table with the information in the first two lines of Table 1 shows that two out of the three components were retained as indicators, containing 20 out of 40 variables that entered in the PCA. These 20 variables are listed on the lines of Table 4 and the third column of the table reports their respective component loadings when the entire dataset is considered. On the other hand, the last two columns of the table show how loadings changed when the PCA was re-run only using the Rome or the Stockholm dataset. When a given variable did not have a significant loading on its corresponding indicator, the table reports whether it had one on the other indicator, or on the third principal component that was not retained as an indicator (“OTHER” label), or if it is not loading on any component at all (“NONE” label). Two italicized lines show the number of additional variables in site-specific analyses not found in the original indicator definition, while the last line reports the sample sizes available for the three PCAs.

Differences in factor loading values on the same row are not so essential to consider, since all loadings above 0.4 are significant. It can therefore be concluded that both indicators RAIL1 and RAIL2 were essentially found again in the Rome dataset: only one variable out of nine in RAIL2 did not have a significant loading in any component, while both indicators from Rome would have three additional variables. Results for the Stockholm dataset show wider differences, since six out of eleven variables are found again in RAIL1 and four out of eight in RAIL2. Even if such differences are certainly not negligible, it seems nevertheless that the relative majority of variables still contribute to the latent construct represented by the two indicators, while no more than two variables would be "moved together" from one indicator to the other. It should also be considered that the Stockholm dataset is quite small to run a PCA with so many variables, so that the results of this specific analysis are less reliable than those for Rome or for the overall dataset.

These results are quite typical of those found for other indicators. When the number of observations available in the test site is sufficiently large, the definition and the structure of the components coming out from a city-specific PCA are comparable with those from the total dataset, since many variables are always grouped together. This is a very important result, since the above proposed indicators can be considered valid, at least to some extent, also beyond the dataset that was used to generate them, despite the fact that PCA is a completely empirical and data-driven exploratory analysis technique. Some variables come to be a sort of common baseline in the composition of the indicators, irrespective of the portion of dataset under consideration, so that the main characterization and the subsequent interpretation of the latent variables are still valid.

## INDICATORS BENCHMARKS AND QUALITY PROFILES OF CITIES

### Computing Indicators Values

The above stability analysis has shown that a benchmarking assessment involving other cities beyond the METPEX test sites can be reasonably made through the above defined indicators. In fact, one should consider that sufficiently large datasets are needed only for the derivation of the factor themselves through Principal Component Analysis. Once the indicators are defined, their mean values can be computed on much smaller samples, although admittedly small samples are still more subject to random sampling variation. For each traveler *j*, the value $I_j$ taken by a generic indicator *I* is the following:

$$I_j = \frac{\sum_{i=1}^{n} C_i V_{i,j}}{\sum_{i=1}^{n} C_i} = \frac{C_1 V_{1,j} + C_2 V_{2,j} + \cdots + C_n V_{n,j}}{C_1 + C_2 + \cdots + C_n}$$

where $C_i$ is the component score coefficient of variable *i* composing the indicator *I* and $V_{i,j}$ is the satisfaction rating, ranging from 1 to 5, given by traveler *j* to item *i*. Then, the average of $I_j$ values for all travelers *j* in a city can be computed. Component score coefficients are an output of previously illustrated PCAs and they should not be confounded with the above mentioned component loadings. In fact, component score coefficients, rather than component loading, should be used to compute the value of a previously defined indicator for new observations, as in the above formula. If **L** is the (*n* variables x *p* components) matrix of loadings and **C** is the *n* x *p* matrix of scores, then $\mathbf{C} = \mathbf{L}(\mathbf{L}^{\mathbf{T}}\mathbf{L})^{\mathbf{-1}}$ when the varimax rotation of factors is applied as in the present research. Component score coefficients of the defined indicators are available in the METPEX indicators manual *(21)*.

Mean values for all 92 indicators in the eight test sites were thus computed through the above formula, to create a benchmarking set of values that can be used as a broad reference. When applying the method, practitioners however need not to use the complete set. The above mentioned manual provides guidance on how to select an appropriate set of indicators according to the evaluation perspective and goals. In turn, the customer satisfaction survey needed to compute the indicators can be much shorter than the one implemented during the project, only asking respondents to rate those items that enter in the indicators definition.

**Indicators dashboard**

The final part of the work involved the definition of city quality profiles as an effective tool to communicate to stakeholders the outcome of the analysis. For each test site, mean values of several different indicators have therefore been plotted together through radar graphs and matched against a benchmark, which is represented by the average over all test sites.

Five different radar graphs have been organized by considering the above multidimensional structure for deriving the indicators. Two graphs respectively report the values of transport mode- and user group-specific indicators that are listed in the first two blocks of Table 1. Two additional graphs are devoted to indicators from Table 2, separately showing them according to the number of components found during the analysis (1 or 2) since there are too many indicators for a single chart. The last graph groups together indicators from the lower tier of Table 1 and those from Table 3.

An example of these five charts is reported in Figure 1. Recalling that the range of values of indicators is from 1 to 5, a plotted value of 0 means that the corresponding indicator could not be computed due to the lack of observations. It can be seen that, jointly considered, these plots constitute a dashboard that can help policy makers to have an overview of how different components of the transport system in a city are evaluated by different user groups, identifying the main areas of improvement and the related measures that should be implemented. By looking at these figures, several insights are possible.

For example, concerning different travel modes (Figure 1*(a)*), the city under consideration seems to perform relatively poorly concerning active means, i.e. pedestrians and bikers. In particular, the mean value of indicator PED2 is not shown due to insufficient sample size, the latter changing according to non-response patterns, while PED3, that is dealing with the "sympathy of design", is the worst compared to the benchmark. Interestingly enough, this indicator is pointing to an issue that is probably not under the control of transport agencies or providers. The METPEX set of indicators tries to keep into consideration that the perceived quality of transport services in urban environments is influenced by a variety of factors and not all of them can be clearly traced back to the main actors of the transport system. On the other hand, public transport and especially bus and metro services are well evaluated.

The chart in Figure 1*(b)* offers the viewpoint of different typologies of travelers. Satisfaction ratings of women and people traveling with children are in line with the average, while those aged under 24 and commuters are particularly satisfied. At the other extreme, communication impaired and visitors are relatively less satisfied. By looking at single indicators, CIMP2 and VISIT1 are dealing with the quality of the information and represent the main areas where an improvement is sought. However, quite intriguingly also YOUNG1 is dealing with information issues, although scores are quite good in this case. This could be due to the fact that the information system is of good quality but relying on the advanced use of information and communication technologies (smartphones, social media etc.), so that is not enough inclusive for individuals that are not familiar with their use. The other main areas of improvement that can be identified are the provisions of services for mobility restricted and elders (indicators MOBR1 and OLD3), cognitive barriers for foreigners (VISIT2) and the comfort for communication impaired (CIMP5). Again, comfort seems an issue only for this specific group of travelers, since other indicators dealing with comfort (such as BUS3, COMM4 and LOW2) have values not below the average.

Quality indicators in the Figure 1*(c)* and *(d)* are quite in line with the average, as opposed to the previous two, with the only exception of QUAL13 (quality of ride). This can be seen as a confirmation of the need to separately looking at different user groups and travel modes to have a better view of the issues at stake. Finally, when considering indicators specific to wait and walking to/from transport service points or traveling onboard transit vehicles (Figure 1*(e)*), it can be confirmed that public transport services in this city are of good quality for the general public. It is on the other hand confirmed that improvements related to the transit systems are desirable for those passengers that are either communication impaired or mobility restricted.

## CONCLUSIONS

The present paper has illustrated a method to build and use a set of quality indicators for transport services that keeps into consideration the disparate views of different groups of individuals. Indicators are based on the output from customer satisfaction surveys, similar to those that are normally carried out to monitor transit services. An innovative aspect of the METPEX project is the investigation of a specific journey experience, rather than general opinions on travelers' attitudes. A potential drawback is that some questions, for example the mode-specific ones, are in principle answered only by people using that specific means for their trip. Therefore, it seems difficult to gather the views of those not using that means with such a tool, albeit it would be of interest. However, we remind that all variables were tagged according to their relevance to a specific evaluation dimension (travel means, traveler type or journey phase), irrespective to where the related question was actually asked. Therefore, thanks to such procedure, all information from a specific questionnaire that is related to a specific mode or user group can be filtered, irrespective of where it was actually asked. For example, many Tier-2 questions are related to public transport, so that even travelers that did not use this means to complete the investigated trip probably expressed their views related to transit in the last part of the survey. All this information is used to define our indicators, thus maximizing the efficiency of the process.

The indicators that were defined derive from the smallest set of latent dimensions of the perceived quality of the journey experience that can explain the maximum amount of the observed variability of ratings. In most cases, the meaning that can be associated to such latent variables contrasts with the immediacy of questions contained in the survey, therefore representing more abstract or less intuitive concepts, for which eliciting judgments through more direct questions would have been less reliable.

The purely empirical approach that was adopted seems to have produced a set of indicators that might not be completely satisfactory from an expert viewpoint. However, any consideration should not be done simply considering the forcedly synthetic indicator names derived in Tables 1, 2 and 3, but rather the lists of satisfaction ratings included in each indicator, that cannot be reported here for space reasons but are available in the METPEX indicators manual *(21)*. It can nevertheless be concluded that some indicators are partially overlapping (for example, indicators with similar but distinct definitions that are capturing the concept of comfort according to the views of different kinds of users). This was totally expected, given the fact that several different parallel analyses of portions of the dataset have been carried out, and it can be useful to better understand differences in the views of individuals on the same issue, that involve not only evaluation aspects but the definition itself of the issue.

The results from the analysis are then visualized through a quality monitoring dashboard. It is believed that this method to build a multidimensional profile of the cities can give on one side a sufficiently broad overview of the perceived quality of the whole transport system, on the other it can help stakeholders in assessing the system components under different viewpoints. Jointly considering indicators from different dimensions of the evaluation exercise (such as travel modes versus users groups) can in fact more precisely bring into light quite specific issues that would be overlooked in a coarser quality monitoring activity.

The use of Principal Component Analyses to define the indicators represents a data-driven methodology, without the need for models whose use is not so common among practitioners. However, the generalizability of results from such empirical approaches is often questionable. Building on the fact that the project fieldwork took place in eight European cities with different characteristics, it was possible to run some preliminary analyses aimed at checking if indicators representing latent components of the variables under consideration changed in the different cities. According to the above findings, the definition and the structure of the factors coming out from a city-specific analysis are in many cases comparable with those from the total dataset. However, a more systematic assessment should be done, possibly involving additional cities. Confirmatory factor analysis and invariance testing could be used to gain additional insights in this issue. Another desirable extension of this work would be the estimation of a measurement model informed by the results of the present analysis, for example through a Structural Equation Modeling technique.

## ACKNOWLEDGEMENTS

The research described in this article was performed within the European project METPEX (A MEasurement Tool to determine the quality of the Passenger Experience, www.metpex.eu), which has received funding from the European Union’s Seventh Framework Programme under grant agreement No. 314354. The analysis summarised in Table 2 was carried out by Alberto Castro.

## REFERENCES

1. de Oña, J., and R. de Oña. Quality of service in public transport based on customer satisfaction surveys: A review and assessment of methodological approaches. *Transportation Science*, Vol. 49, No. 3, 2015, pp. 605-622. http://dx.doi.org/10.1287/trsc.2014.0544.
2. Transit Cooperative Research Program. *Transit Capacity and Quality of Service Manual*, TCRP Report 165, Third Edition. Transportation Research Board, Washington, D.C., 2013.
3. Transit Cooperative Research Program. *A Guidebook for Developing a Transit Performance-Measurement System*, TCRP Report 88. Transportation Research Board, Washington, D.C., 2003.
4. European Standard EN 13816. *Transportation - Logistics and services - Public passenger transport - Service quality definition, targeting and measurement*. Brussels, European Committee for Standardization, 2002.
5. Lu, A., S. Aievoli, J. Ackroyd, C. Carlin, and A. Reddy. Passenger environment survey: representing the customer perspective in quality control. *Transportation Research Record: Journal of the Transportation Research Board,* No. 2112, 2009, pp. 93-103. http://dx.doi.org/10.3141/2112-12.
6. Carreira, R., L. Patrício, R.N. Jorge, C. Magee, and Q. V. Eikema Hommes. Towards a holistic approach to the travel experience: A qualitative study of bus transportation. *Transport Policy*, Vol. 25, 2013, pp. 233-243. http://dx.doi.org/10.1016/j.tranpol.2012.11.009.
7. Carreira, R., L. Patrício, R.N. Jorge, and C. Magee. Understanding the travel experience and its impact on attitudes, emotions and loyalty towards the transportation provider–A quantitative study with mid-distance bus trips. *Transport Policy*, Vol. 31, 2014, pp. 35-46. http://dx.doi.org/10.1016/j.tranpol.2013.11.006.
8. Tyrinopoulos, Y., and C. Antoniou. Public transit user satisfaction: Variability and policy implications. *Transport Policy*, Vol. 15, No. 4, 2008, pp. 260-272. http://dx.doi.org/10.1016/j.tranpol.2008.06.002.
9. Susilo, Y. O., and O. Cats. Exploring key determinants of travel satisfaction for multi-modal trips by different traveler groups. *Transportation Research Part A: Policy and Practice*, Vol. 67, 2014, pp. 366-380. http://dx.doi.org/10.1016/j.tra.2014.08.002.
10. Andreassen, T. W. (Dis)satisfaction with public services: The case of public transportation. *Journal of Services Marketing*, Vol. 9, No. 5, 1995, pp. 30-41. http://dx.doi.org/10.1108/08876049510100290.
11. Dell’Olio, L., A. Ibeas, and P. Cecín. Modelling user perception of bus transit quality. *Transport Policy*, Vol. 17, No. 6, 2010, pp. 388-397. http://dx.doi.org/10.1016/j.tranpol.2010.04.006.
12. Eboli, L., and G. Mazzulla. A new customer satisfaction index for evaluating transit service quality. *Journal of Public Transportation*, Vol. 12, No. 3, 2009, pp. 21-37. http://dx.doi.org/10.5038/2375-0901.12.3.2.
13. Rojo Arce, M., H. Gonzalo-Orden, L. Dell'Olio, and Á. Ibeas Portilla. Modelling gender perception of quality in interurban bus services. *Proceedings of the Institution of Civil Engineers - Transport*, Vol. 164, Issue TR1, 2011, pp. 43-53. http://dx.doi.org/10.1680/tran.9.00031.
14. Bordagaray, M., L. Dell'Olio, A. Ibeas, and P. Cecín. Modelling user perception of bus transit quality considering user and service heterogeneity. *Transportmetrica A: Transport Science*, Vol. 10, No. 8, 2014, pp. 705-721. http://dx.doi.org/10.1080/23249935.2013.823579.

15. Susilo, Y. O., R. Abenoza, R., A. Woodcock, F. Liotopoulos, A. Duarte, J. Osmond, A. Georgiadis, G.R. Hrin, P. Bellver, F Fornari, V. Tolio, E. O'Connell, I. Markucevičiūtė, and M. Diana. Findings from measuring door-to-door travellers' travel satisfaction with traditional and smartphone app survey methods in eight European cities. Paper presented at the *International Conference on Travel Behaviour Research (IATBR 2015)*, July 19-23 2015, Windsor, UK.
16. Diana, M., M. Pirra, A. Castro, A. Duarte, V. Brangeon, C. Di Majo, D. Herrero, G.R. Hrin, and A. Woodcock. Development of an integrated set of indicators to measure the quality of the whole traveller experience. *Transport Research Procedia*, Vol. 14, 2016, pp. 1164-1173. http://dx.doi.org/10.1016/j.trpro.2016.05.187.
17. Cronbach, L. Coefficient alpha and the internal structure of tests. *Psychometrika*, Vol. 16, 1951, pp. 297-334. http://dx.doi.org/10.1007/BF02310555.
18. Bishop, C. M. *Pattern Recognition and Machine Learning*. Springer, New York, 2006.
19. *A comprehensive set of quality and accessibility indicators for transport services*. Deliverable 5.2 of the METPEX project, 2015.
20. *Testing the indicators and developing benchmarks*. Deliverable 5.3 of the METPEX project, 2015.
21. *A transport services quality and accessibility evaluation manual*. Deliverable 5.4 of the METPEX project, 2015.

**LIST OF TABLES**



**LIST OF FIGURES**

**TABLE 1 Evaluation Dimensions, Tags of Variables and Subsequent Components and Indicators**

| Evaluation Dimension | Tag | No. Var. | No. Comp. | Final List of Indicators |
|---|---|---|---|---|
| Travel mode | Railway services | 40 | 3 | RAIL1: On-trip performance |
| | | | | RAIL2: Ticketing and performances before boarding |
| | Underground | 38 | 3 | UNDER1: On-trip performance |
| | | | | UNDER2: Ticketing and capillarity |
| | Tramways | 37 | 6 | TRAM1: Service integration and reliability |
| | | | | TRAM2: Comfort and staff helpfulness |
| | | | | TRAM3: On-trip quality |
| | | | | TRAM4: Tickets and timetabling |
| | | | | TRAM5: Fares convenience |
| | | | | TRAM6: Information quality |
| | Bus services | 38 | 4 | BUS1: Reliability |
| | | | | BUS2: Ticketing and other issues |
| | | | | BUS3: Comfort on board |
| | Private vehicles | 24 | 2 | CAR: Traffic calming and parking |
| | Pedestrians | 49 | 7 | PED1: Information and safety |
| | | | | PED2: Environmental aspects |
| | | | | PED3: Sympathy of design |
| | | | | PED4: Intermodal travel and barrier free accesses |
| | | | | PED5: Pavements cleanliness |
| | Bikers | 57 | 9 | BIKE1: Cycling infrastructure |
| | | | | BIKE2: Quality of Information, availability and infrastructure |
| | | | | BIKE4: Public transport plus bike trip quality |
| | | | | BIKE5: Easiness in carrying bicycles |
| User group | Women | 48 | 4 | WOM1: Safety and security, comfort and staff helpfulness |
| | | | | WOM2: Integrated tickets and range of fares |
| | | | | WOM3: Reliability |
| | Commuters | 50 | 5 | COMM1: Facilities and parking |
| | | | | COMM2: Reliability |
| | | | | COMM3: Ticketing issues |
| | | | | COMM4: Comfort aspects |
| | Travelers over 65 | 34 | 4 | OLD1: Information and accessibility |
| | | | | OLD2: Performance issues for elders |
| | | | | OLD3: Travel services |
| | | | | OLD4: Infrastructural design |
| | Travelers under 24 | 43 | 4 | YOUNG1: Service design and information |
| | | | | YOUNG2: Reliability |
| | Low income travelers | 40 | 4 | LOW1: Low cost services issues |
| | | | | LOW2: Comfort |
| | Visitors | 36 | 3 | VISIT1: Information issues |
| | | | | VISIT2: Barriers for foreigners |
| | Rural dwellers | 41 | 4 | RURAL1: Service coverage and costs |
| | | | | RURAL2: Ground aspects |
| | | | | RURAL3: Comfort and availability |
| | Traveling with children | 34 | 5 | CHILD1: Design and quality on-board |
| | | | | CHILD2: Ground aspects |
| | Mobility restricted | 39 | 4 | MOBR1: Quality issues specific for mobility restricted |
| | | | | MOBR3: Ability to meet individual needs |
| | | | | MOBR4: Ground services |
| | Communication impaired | 87 | 10 | CIMP1: Quality issues specific for communication impaired |

| Evaluation Dimension | Tag | No. Var. | No. Comp. | Final List of Indicators |
|---|---|---|---|---|
| | | | | CIMP2: Information issues specific for communication impaired<br>CIMP4: On-board quality and information<br>CIMP5: Comfort<br>CIMP6: Information aspects |
| Journey phase | Traveling on-board public transport means | 29 | 4 | ON-PT1: Staff behavior<br>ON-PT3: Trip-specific aspects<br>ON-PT4: Service operations characteristics |
| | Trav. on private means | 26 | 2 | ON-PR2: Trip-specific aspects |
| | Wait and walk to/from a transit service point | 7 | 3 | PT-WW1: Infrastructures design<br>PT-WW2: Trip-specific aspects<br>PT-WW3: Safety and security |

**TABLE 2 Tier-1 Variables and Subsequent Components and Indicators Focusing on Quality Aspects**

| Tier-1 Variable | No. Var. | No. Comp. | Final List of Indicators |
|---|---|---|---|
| 1. Design of stations was adequate for my needs | 12 | 1 | QUAL1: Design of transport stations |
| 2. Design of transport interchanges (main terminals) was efficient | 13 | 1 | QUAL2: Design of transport interchanges |
| 3. Design of transport stops was adequate for my needs | 13 | 1 | QUAL3: Design of transport stops |
| 4. The city supported my mobility needs | 14 | 1 | - |
| 5. The different modes of transport I used worked well together | 14 | 2 | QUAL5A: Quality of crossings<br>QUAL5B: Physical interactions among modes |
| 6. My passenger rights (e.g. able to access all transport services) were respected | 14 | 2 | QUAL6B: Non-discriminatory service and protection of data |
| 7. The overall accessibility of my journey was adequate for travelers with additional needs | 14 | 2 | QUAL7A: Accessibility for different user groups<br>QUAL7B: Social dimension of services |
| 8. Provision of information on arrivals and departures was adequate for my needs | 10 | 1 | QUAL8: Provision of information on arrivals and departures |
| 9. Public Transport Staff were receptive to my needs | 11 | 1 | QUAL9: Public transport staff |
| 10. The quality of travel information available during journey was good | 17 | 1 | QUAL10: Quality of travel information during journey |
| 11. The quality of pre-trip information before I started my journey was good | 14 | 1 | QUAL11: Quality of pre-trip information before the journey |
| 12. The quality of transport infrastructure during my journey was good | 13 | 2 | QUAL12A: Overall quality of transport infrastructure<br>QUAL12B: Vandalism and graffiti |
| 13. The quality of my ride was good | 7 | 1 | QUAL13: Quality of ride |
| 14. My safety and security while traveling was good | 12 | 1 | QUAL14: Safety and security while travelling |
| 15. Support for intermodal (different forms of transport during the journey) travel was provided | 13 | 1 | QUAL15: Support for intermodal travels |
| 16. Recognition of needs of motorized vehicle users | 12 | 1 | QUAL16: Motorized vehicle users' needs |
| 17. Ticket purchasing process was easy to follow | 14 | 2 | QUAL17A: Tickets regulations and flexibility<br>QUAL17B: Practical aspects related to ticketing |
| 18. Time the journey took was as promised | 12 | 1 | QUAL18: Reliability and on-time performance |
| 19. Transport availability was adequate for my needs | 13 | 1 | QUAL19: Service availability |
| 20. Vehicle design was suitable for my needs | 13 | 2 | QUAL20A: Ergonomy<br>QUAL20B: Design for specific user groups |
| 21. Value for money of services was good | 13 | 1 | QUAL21: Value for money of services |

*- = indicator not defined*

**TABLE 3 Cross-Classifications Analyses and Subsequent Components and Indicators Focusing on Mobility Challenged Individuals Using Public Transport**

| Mobility Impairment | No. Var. | No. Comp. | Final List of Indicators |
|---|---|---|---|
| Mobility restricted | 100 | 8 | PT-MR1: Infrastructural design<br>PT-MR2: Relevant features for mobility restricted<br>PT-MR3: Service operations<br>PT-MR4: Ticketing and other issues |
| Communication impaired | 94 | 8 | PT-CI1: Communication aspects<br>PT-CI2: Information design<br>PT-CI3: Punctuality and reliability |

## TABLE 4 Significant Component Loadings in the Whole Dataset and in Two Cities

| Indicator | Variables | Overall Loading | Rome Loading | Stockholm Loading |
|---|---|---|---|---|
| RAIL1 | Level of noise | 0.755 | 0.735 | 0.475 |
| | Level of crowding | 0.747 | 0.621 | 0.649 |
| | Punctuality | 0.661 | 0.620 | 0.779 |
| | Notification on timetabling changes | 0.651 | 0.483 | 0.507 |
| | Cleanliness of vehicles | 0.613 | 0.689 | OTHER |
| | Reliability at off peak times | 0.603 | 0.631 | 0.520 |
| | Reliability of services | 0.594 | 0.492 | 0.557 |
| | Helpfulness of customer facing staff | 0.580 | 0.564 | RAIL2 |
| | Frequency of services | 0.519 | 0.527 | RAIL2 |
| | Safety and security at transport stops | 0.518 | 0.734 | NONE |
| | Air temperature and ventilation inside vehicles | 0.518 | 0.527 | OTHER |
| | *Number of additional variables in site-specific analyses* | - | *3* | *4* |
| RAIL2 | Ability to buy one ticket which covers different forms of transport | 0.652 | 0.621 | 0.473 |
| | Range of fares offered | 0.674 | 0.497 | 0.805 |
| | Comprehensibility of ticketing structure | 0.644 | 0.585 | 0.618 |
| | Accessibility of platforms | 0.590 | 0.725 | RAIL1 |
| | Availability of ticket buying locations | 0.589 | 0.576 | RAIL1 |
| | Value for money of services was good | 0.581 | 0.475 | 0.604 |
| | Accessibility of station facilities | 0.560 | 0.642 | OTHER |
| | Ticket purchasing process was easy to follow | 0.543 | 0.650 | 0.434 |
| | Distance from Origin or Destination to closest station | 0.455 | NONE | OTHER |
| | *Number of additional variables in site-specific analyses* | - | *3* | *4* |
| *Sample size* | | *918* | *224* | *131* |

*- = data not applicable*

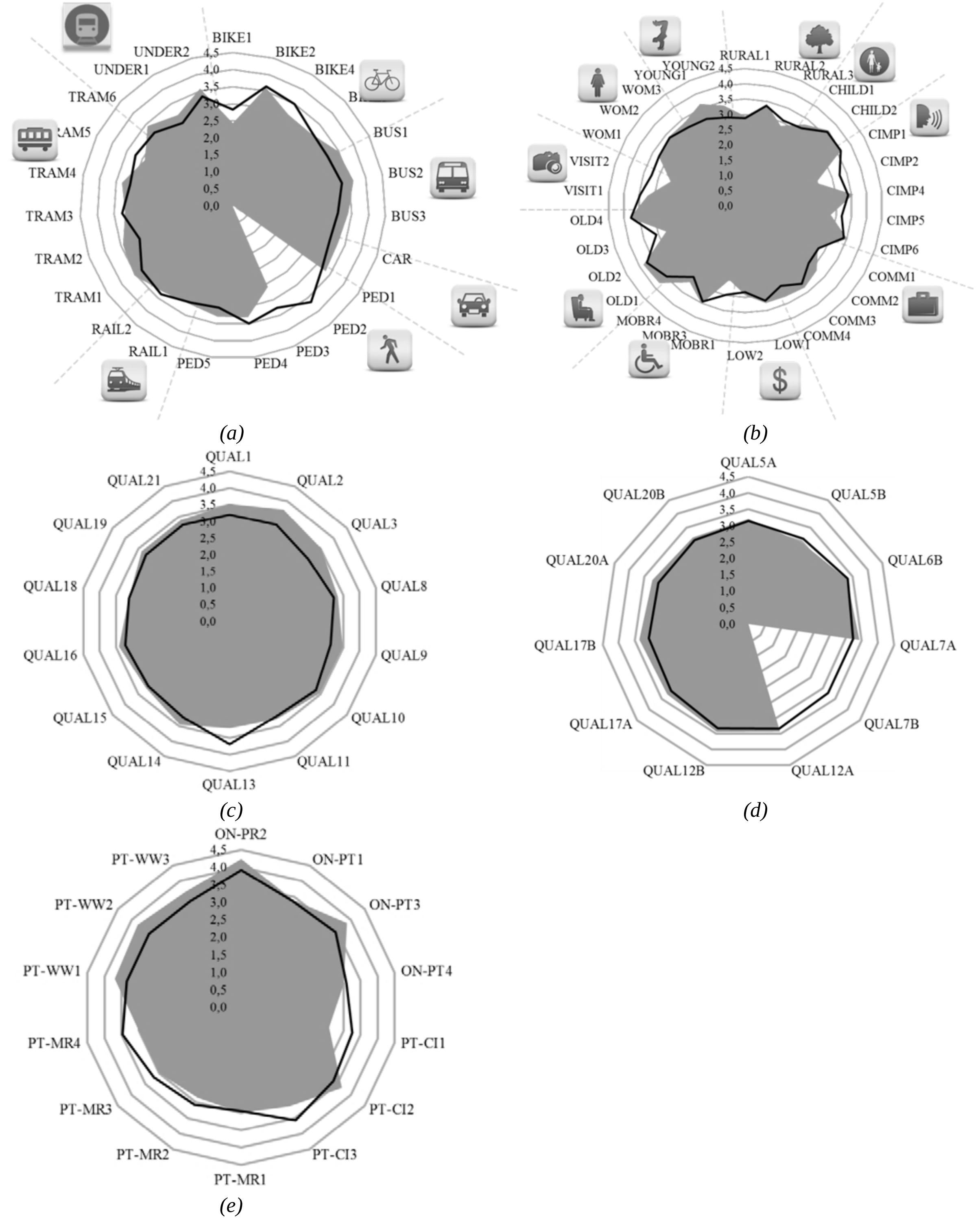


**FIGURE 1 Transport quality profile of a city (grey polygon) against the benchmark (black polyline) for mode indicators *(a)*, user groups indicators *(b)*, quality indicators with one *(c)* or two components *(d)*, remaining indicators *(e)*. All features are described in Table 1, 2 and 3.**