EMPIRICAL　　　　　　　　　　　　　　　　　　　　　　**Open Access**

Customer service quality and benchmarking in public transport contracts

David A Hensher

Correspondence:
David.Hensher@sydney.edu.au
Institute of Transport and Logistics
Studies (ITLS), The University of
Sydney Business School, The
University of Sydney, Sydney, NSW
2006, Australia

Abstract

As contracting of public transport services increases in sophistication, there is a growing focus on an increasing number of key performance indicators that emphasis service quality. Although contracts won under competitive tendering or by negotiation are assessed on a number of evaluation criteria, cost efficiency still remains the main basis for selecting a preferred operator. There has been a limited effort to identify the service quality influences that really matter to users of public transport. Ways of incorporating the packaging of service quality offer an improved and behaviourally richer way of representing the role of underlying dimensions of quality in establishing how well a contracted service is delivering services to satisfy customers. In this paper we present a way of doing this using a construct called a Customer Service Quality Index (CSQI), in which a stated preference survey together with actual experience in using public transport is used to obtain preference weights for each significant attribute defining service quality, and which is used then to establish a CSQI for each sampled user, and by aggregation, the performance on service quality of each operator. Such a measure should be considered by regulators when both assessing the merits of each operator's bid in order to avoid the real risk that cost efficiency dominates at the expense of gains in service performance and in ongoing monitoring of performance.

Keywords: Customer service quality; Benchmarking; Bus contracts; Stated preference surveys; Logit models

Background

As contracting of public transport services grows in sophistication, there is a growing focus on an increasing number of key performance indicators that emphasis service quality. Although contracts won under competitive tendering or by negotiation are often assessed on a number of evaluation criteria, cost efficiency still remains the main basis for selecting a preferred operator. The reasons are in part linked to the need to reduce costs through competition for the market (in tendering), in contrast to other models of contract awarding. In both tendered and negotiated contracts, the ideals of performance improvement are promoted, yet it is often the case that this is poorly understood, and where it is taken as a precise measure of output it is limited to a few measurable criteria that often are defined by supply side measures (e.g., on-time running and accidents per 100,000 operating hours).

There has been a limited effort to identify the influences that really matter to users of public transport; however the efforts have been based on simplistic measures of

satisfaction measured of a Likert scale (typically from very unsatisfied to very satisfied), where the contributing influences are treated as independent and hence additive in their impact. In reality, users of public transport purchase a package of attributes that define perceptually what matters to them. Ways of incorporating such packaging of service quality offer an improved and behaviourally richer way of representing the role of underlying dimensions of quality in establishing how well an operator is delivering services to satisfy customers and also their contract obligations. Studies, for example, by Cirillo *et al.* [1], dell'Olio *et al.* [2], Eboli and Mazzulla [3] have been critical of methods that treat each underlying service dimension as if it can stand alone in the way it is assessed as an influence on public transport performance (or indeed for any performance of any market delivered service). Bolton and Drew [4] and Boulding *et al.* [5] are typical examples of customer satisfaction studies that promote attribute packaging in determining the value of service.

In this paper we present a novel way of doing this using a construct developed by Hensher some years ago called a Service Quality Index (SQI) (see [6]), in which a stated preference survey together with actual experience in using public transport is used to obtain preference weights for each significant attribute defining service quality, and which is used then to establish an SQI for each sampled user and by aggregation the performance on service quality of each operator. Such a measure should be considered by regulators when assessing the merits of each operator's bid in order to avoid the real risk that cost efficiency dominates at the expense of gains in service performance.

The focus of this paper is quantifying service quality from a users' perspective, in a way that weights the relevance of each source of service quality, and to use this information to obtain a single measure of service quality, that we call the customer service quality index. We use data from metropolitan and non-metropolitan bus operators in New South Wales (NSW) to demonstrate the way the method can be used in contract negotiation and ex-post monitoring of performance leading up to contract renegotiation or tendering.

Concerns about traditional Likert scale metrics of customer satisfaction

Service quality in the context of customer experience is typically identified using a Likert scale in which consumers are asked to indicate on a scale (such as from very satisfied to very unsatisfied), how satisfied they are with a specific attribute defining a class of service. Sometimes these satisfaction measures are weighted by a reported importance attached to an attribute, often referred to as the Fishbein-Ajzan importance-satisfaction scale. This approach falls short in two major ways of creating a robust customer service quality index. First, there are interpretations of scale issues with surveys which rely on respondents marking off the different aspects of service quality. We do know that 'very satisfied' to one user might be rated as 'very dissatisfied' from another user on exactly the same service. Second, there are independent issues with the way in which respondents are asked to consider the different aspects of service quality. The literature [7] is clear that users do not find all aspects of service quality equally *important* and, indeed, may be extremely satisfied with an element which is of low importance.

Within the broader context of SERVQUAL that has had a dominating influence in consumer and marketing research, Buttle [8]) provides an extensive review and critique

of the SERVQUAL method first introduced by Parasuraman *et al.* [9]. The approach using ordinal scaled (Likert) metrics performs analyses with methods suited to interval-level data (factor analysis) [10]. It has been criticised on many grounds including that interdependencies among the dimensions of quality are difficult to describe. Importantly it has been criticised for focusing on the process of service delivery rather than outcomes of the service encounter (Richard and Allaway [11].

For this reason, the best approach to the customer service index would be adopt a customer service quality index (CSQI) approach in which an methods such as a stated preference (SP) experiment investigates with users their response to variations around the current level of service attributes, presented in packages (as illustrated in the case study in a later section), which are created using formal statistical design principles (see Hensher *et al.* [12], 2005). The data would be modelled by combining the SP data from the experiment and the revealed preference (RP) data on currently perceived levels of experience in respect of the attributes of interest from a survey, to obtain estimated parameters for each service quality attribute, which can then be combined to create a customer service quality index.

These weights would not be subject to the interpretation of scale issues identified above and a CSQI thus designed would meet the requirements of being robust, capable of benchmarking customer satisfaction over time to inform policy and planning of services, and to inform how the separate elements of service quality contribute to service performance with a view to informing contract management.

The CSQI created is critically dependent on the questions asked of users. In turn, this requires clarity in the potential use of the customer service quality index. For example, in the context of a bus contract, the interest herein, a CSQI which is to be used for monitoring consumer sentiment and for evaluating policy and planning through the impact of such policies and planning on the service quality attributes, can be less specific than a CSQI which is also to be used for identifying which elements of service quality contribute to performance for the purposes of operator contract development, benchmarking and compliance. In the latter case, the level of service service quality attributes need to be carefully constructed to create, so far as possible, attributes where the responsibility for changes to the level of service can be attributed to a single stakeholder (operator or government). Moreover, if the CSQI is to be used for operator contract development, benchmarking and compliance, a larger sample of bus users in particular will be required as it will be important to capture data effectively for each contract area.

In developing a CSQI for use in benchmarking of operator performance, it is critical to recognise that operator's may have little control over many of the attributes that define a consumer service quality index (or any customer satisfaction metric). Hensher, D.A. [13] has shown this in a study of private bus operators in Sydney (Australia) and concluded that many of the service quality dimensions that matter to users of public transport are often defined in a contract by the regulator (e.g., service coverage by time of day and day of the week and weekend, service frequency, while some attributes are heavily influenced by market forces (e.g., travel times at the peril of traffic congestion and weather).

This paper focuses on the role that customer satisfaction expressed as a CSQI can play in the benchmarking of operator performance as part of monitoring bus contract

outcomes by the funder. Whilst this paper provides evidence for the bus industry, there is no reason to believe that the conclusions would differ for other public transport modes. In Hensher [13] we investigated the relationship between cost efficiency and CSQI elements under the control of the operator, and concluded that operators who provide services levels that result is a higher CSQI in general are also the more cost efficient, and that investing in higher levels of customer service does not have to necessarily require a great cost outlay. In the current paper we are focussing only on the development of a CSQI, providing details of how this can be obtained and used in a benchmarking context for regulators interested in both annual reviews of incumbent operators as well as for competitive tendering or negotiation (and re-negotiation) on contracts. The inclusion of CSQI in the performance review means that benchmarking becomes are much more valued process than simply one based on cost efficiency which is commonly mistaken as the same as cost reduction or no cost escalation (beyond inflation and other agreed variations).

Developing a customer service quality index

The concept of customer service quality includes aspects of transport service which are not always well-defined and easily measured. Herein we define service quality in terms of a set of attributes which each user perceives to be the sources of utility (or satisfaction) in bus use. The dimensions of quality, viewed from a bus user's perspective, are complex. Passengers might, for example, consider the comfort at the bus stop and the time to get a seat, or only the comfort of the seats. Modal choice surveys have identified a large number of influences on the use of buses in contrast to other private and public modes.

Service quality can be divided into six broad classes of effects, summarised in Table 1, each containing different quality dimensions (as identified by [14-18], and other studies). Recent contributions by Cirillo et al. [1], dell'Olio et al. [19], Eboli and Mazzulla[a] [3,20,21] and Marcucci and Gatta [22] have also reinforced the relevance of the attribute set identified in earlier research. Some of these contributions also use a stated preference method, acknowledging the original contribution by Hensher (estimated as multinomial logit and mixed logit models– see [6]), while other studies use a different, more traditional method, in which a satisfaction scale is multiplied by an importance scale (in various forms) to obtain an overall customer satisfaction index. The importance weights are used as proxies for the weights obtained from model estimation herein.

Some demand side measures can be translated (or mapped) into a set of supply side equivalences (resources that the operator has partial or total control of) such as the timetable, fleet age, and/or the buses that are air conditioned; the number of vehicles that are wheelchair accessible, the number of cleaning hours of the vehicles, and the money spent on driver training.

The attributes on the supply side are, in contrast to the quality attributes in column two in Table 1, to varying degrees, observable and under the direct control of the bus operator. For example a change in the average fleet size will, *ceteris paribus,* have a direct influence on the time to get a seat. On the other side we expect the supplied level of service quality to be a function of consumer preferences. If the supplied quality level is a response to customer preferences, and not only to some regulatory restrictions,

Table 1 Demand side effects and their equivalence on the supply side

	Dimensions of user perceptions	Operator specified measures
Getting to the bus stop quality	ease, safe, time (distance), knowing where the bus stop is	frequency, availability of bus shelter and seats
Wait quality	wait time at stop, punctuality of bus	frequency
	wait comfort, wait safety	availability of bus shelter and seats
Trip quality	time to board a bus	frequency, % of low floor buses
	time to get a seat	number of seats available
	moving to your seat	average speed, network shape
	travel time	travel time
	trip cost	fare
Vehicle quality	cleanliness	hours of vehicle cleaning/vehicle
	comfort of seats (types), spaciousness	percent of buses with cloth seats
	temperature control (ventilation)	percent of buses with air conditioning
	noise	visual surveillance
	safety	average age of the fleet
	modernity	wheelchair access (yes/no)
	ease of use for those with disabilities	
Driver quality	appearance-helpfulness	years of driving experience
		money spent on driver training
Information quality	pre-trip information	availability of timetable/destination signs

quality exogeneity cannot be assumed. In this circumstance we need to develop a capability to represent the quality of service as determined by users. The discrete choice approach is an appealing framework (see below). Given these considerations about service quality, we are able to introduce an improved version of the traditional cost model in its reduced form to capture the full dimensionality and service quality.

The proposed and preferred service quality measure is constructed by analysing bus user preferences for different levels of bus service quality, and using the resulting weights attached to each underlying dimension of service quaility as perceived by users to derive the level of satisfaction associated with the supplied level of service quality. To this extent we need to identify and quantify the preferences for service levels from bus travellers. We restrict our analysis to actual bus users but recognise that non-users also provide useful information on the levels of service offered by bus operators. Within a performance regime based on the acceptability of service levels to actual users, and with a focus on the service quality that influences operator costs, the emphasis on users is appropriate.

To reveal user preferences for service quality, we need to obtain data of sufficient richness to capture the behavioural responses to a wide range of levels of service quality defined on an extended set of attributes such as those given in Table 1. Revealed preference (RP) data is typically restrictive in its variance properties, but is an important input into the assessment. The preferred approach is a stated preference (SP) experiment combined with perceptions of existing levels of service. A sampled passenger would evaluate a number of alternative service levels (known as scenarios) together with the level experienced, and choose the most preferred alternative. Systematically varying the levels of the attributes in repeated scenarios enables us to obtain a profile of each passengers

preferences for bus services. The data is analysed as a discrete choice model in which we combine the SP and RP data to obtain estimated parameters for each attribute.

We estimate the simple multinomial logit model (MNL) in which all random components are independently and identically distributed (IID) (see Hensher *et al.* [12], 2005)[b]. Let U_{nsj} denote the utility of alternative j perceived by respondent n in choice situation s. We assume that U_{nsj} may be partitioned into two separate components, an observable component of utility, V_{nsj} and a residual, unobservable component, ε_{nsj}, such that

$$U_{nsj} = V_{nsj} + \varepsilon_{nsj}. \tag{1}$$

The observable component of utility is typically assumed to be a linear relationship of observed attribute levels, x, of each alternative j and their corresponding weights (parameters), β, with a positive scale factor, σ_n such that

$$U_{nsj} = \sigma_n \sum_{k=1}^{K} \beta_{nk} x_{nsjk} + \varepsilon_{nsj}, \tag{2}$$

where β_{nk} represents the marginal utility or parameter weight associated with attribute k for respondent n. The unobserved component, ε_{nsj}, is often assumed to be independently and identically (IID) extreme value type 1 (EV1) distribution. We will develop the implications of the distributional assumption in detail below. The individual scale factor in Equation (2) is normalised to one in most applications. (We refer to such models as *constant variance models*.) An alternative representation that preserves the preference order in Equation (2), as long as σ_n does not vary across alternatives, is

$$U_{nsj}^* = \sum_{k=1}^{K} \beta_{nk} x_{nsjk} + \left(\varepsilon_{nsj}/\sigma_n\right). \tag{3}$$

It can be seen that the variance of ε_{nsj}, is inversely related to the magnitude of $\sigma_n \sum_{k=1}^{K} \beta_{nk} x_{nsjk}$ via σ_n. If ε_{nsj} has an EV1 distribution with this scale parameter, then $\mathrm{Var}(\varepsilon_{nsj}/\sigma_n) = \pi^2/6$. In order to make any progress at modelling choices, it is necessary to make a number of assumptions about the unobserved components of utility. The most common assumption is that for each alternative, j, ε_{nsj}, will be randomly distributed with some density, $f(\varepsilon_{nsj})$, over decision makers, n, and choice situations, s. Further assumptions about the specific density specification adopted for the unobserved effects, ε_{nsj} (e.g., the unobserved effects are drawn from a multivariate normal distribution) lead to alternate econometric models.

Assuming there exists some joint density such that $\boldsymbol{\varepsilon}_{ns} = \langle \varepsilon_{ns1}, ..., \varepsilon_{nsJ} \rangle$ represents a vector of the J unobserved effects for the full choice set, it becomes possible to make probabilistic statements about the choices made by the decision makers. Specifically, the probability that respondent n in choice situation s will select alternative j is given as the probability that outcome j will have the maximum utility;

$$\begin{aligned} P_{nsj} &= \mathrm{Prob}\left(U_{nsj} > U_{nsi}, \forall i \neq j\right) \\ &= \mathrm{Prob}\left(V_{nsj} + \varepsilon_{nsj} > V_{nsi} + \varepsilon_{nsi}, \forall i \neq j\right) \end{aligned} \tag{4}$$

which can also be written as

$$P_{nsj} = \mathrm{Prob}\left(\varepsilon_{nsj} - \varepsilon_{nsi} > V_{nsi} - V_{nsj}, \forall i \neq j\right). \tag{5}$$

Equation (5) reflects the probability that the differences in the random terms, $\varepsilon_{nsi} - \varepsilon_{nsj}$ will be less than the differences in the observed components of utility, $V_{nsi} - V_{nsj}$. The probabilities for a multinomial logit model given equation (5) can be computed in closed form. It has been shown in many sources, such as [12,23], Train, [24]), that for a multinomial logit (MNL) model:

$$\text{Prob(Alt j is chosen)} \;=\; \frac{\exp\left(V_{nsj}\right)}{\sum_{j=1}^{J}\exp\left(V_{nsj}\right)}, j = 1, ..., J. \tag{6}$$

Assuming that the utility functions themselves are straightforward, the probabilities in (6) can be computed simply by plugging relevant quantities into the formula, with no approximations required. This is one of the appealing features of the multinomial logit form of a choice model which we use in this study.

A CSQI for each bus operator (or contract region) can be derived from the application of the parameter estimates obtained form the estimation of the MNL model to the current RP levels which each operator-specific passenger sample currently experiences. This index is not a probability (of choice) weighted indicator that is typically derived from a choice model; rather we seek to establish an indicator based solely on the levels of service currently on offer. The SP-RP model's role is to provide a rich set of parameter estimates to weight each attribute of service quality.

To assist in the selection of attributes for the CSQI, we undertook an extensive review of the literature as well as a survey of bus operators who have a wealth of experience on what customers look for in a good service (see [25]). We found that thirteen attributes describe the major dimensions of service quality from a user's perspective. The range of levels of each attribute in Table 2 provided us with a mechanism for establishing the weights that signal the contribution of each attribute to the overall SQI.

Through a formal statistical design, the attribute levels are combined into *bus packages* before being translated into a survey form. The full factorial design (i.e., all possible bus packages) consists of 3^{13} combinations of the 13 attributes each of three levels. To produce a practicable and understandable design for the respondents, we restricted the number of combinations to 81 (i.e., 81 choice sets) using a fractional design. Fractional designs permit the reduction in the number of combinations (i.e., the number of bus packages) without losing important statistical information (see [23]).

A pre-test of the survey showed that respondents were able to evaluate consistently three choice sets (i.e., different scenarios of bus packages), resulting in 27 different survey forms. To allow for a rich variation in the combinations of attribute levels to be evaluated as service packages in the SP experiment, each bus operator received 8 sets of 27 different survey forms (i.e., 216 forms) and instructions on how to organise the survey. An example of an SP question is shown in Additional file 1: Table S3a, with the questions on a recent trip and background data shown in Additional file 1: Table S3b.

Scheduled[c] bus users of 25 private bus operators in NSW participated. Survey forms were distributed and collected during the first half of 1999. A total of 3,849 useable observations (out of 4,334 returns) were incorporated in the estimation of the discrete choice model. A multinomial logit (MNL) specification was selected. This is appropriate for a model form in which the utility expressions associated with the current trip

Table 2 Set of attributes and attributes levels in the SP experiment

Attribute	Interpretation of levels	Attribute	Interpretation of levels
Reliability	on time	**Info at the bus stop**	on time
	5 minutes late		5 minutes late
	10 minutes late		10 minutes late
Frequency	every 15 minutes	**Travel time**	25% quicker than the current travel time
	every 30 minutes		same as now
	every 60 minutes		25% longer than the current travel time
Walking distance to the bus stop	-same as -now	**Bus stop facilities**	bus shelter with seats
	5 minutes more		seats only
	10 minutes more		no shelter or seats at all
Waiting safety	very safe	**Fare**	25% more than the current one-way fare
	reasonably safe		same as now
	reasonably unsafe		25% less than the current one-way fare
Access to the bus	wide entry with no steps	**Driver attitude**	very friendly
	wide entry with 2 steps		friendly enough
	narrow entry with 4 steps		very unfriendly
Air conditioning	available with no surcharge	**Safety on board**	the ride is very smooth with no sudden braking
	available with a surcharge of 20% on existing one-way fare		the ride is generally smooth with rare sudden braking
	not available		
Cleanliness of seats	very clean		the ride is jerky; sudden braking occurs often
	clean enough		
	not clean enough		

and two attribute packages are unlabelled (or unranked) alternatives. Consequently all design attributes were generic across the three alternatives. In addition, in the current trip alternative we considered alternative-specific characteristics of the passenger (income, gender, age and car availability) and of the operator together with a number of other potential influences on relative utility such as treatment effect, trip purpose and access mode.

The user preference model results

The user attribute choice model is summarised in Table 3 9 with acronyms defined in the Appendix. The model includes the attributes of the SP experiment, operator-specific dummy variables and three user characteristics. The nine sets of two dummy variables per service attribute are defined relative to a third level which is set to zero (given the three levels in the design). The overall goodness of fit (adjusted pseudo-R^2) of the model is 0.324. The great majority of the design attributes are statistically significant. Service reliability (i.e., the extent to which buses arrive on time), fares, access time and travel time are all highly significant with the expected negative sign. Relative to 'reasonably unsafe', we find a positive (almost) significant parameter estimate for 'reasonably safe' (0.1510) and for 'very safe' (0.1889). The higher estimate for 'very safe' in contrast to 'reasonably safe' is plausible. The infrastructure at the bus stop appears not

Table 3 Final user preference model

Variable	Units	Acronym	Parameter	t-value
Reliability (all levels)	mins	RELI	-.05821 -	8.411
Bus one-way fare (all levels)	$	TARIF	-.4780 -	6.406
Access walk time (all levels)	mins	ACCESST	-.04317 -	5.311
Bus time (all levels)	mins	TRATIM	-.03200 -	5.435
Personal safety at bus stop:				
Very safe	1,0	VSAFE	.18895	2.255
Reasonably safe	1,0	RSAFE	.15108	1.820
Bus stop facilities:				
Seats only at bus stop	1,0	SEATS	-.03411	.510
Seat plus shelter	1,0	SEATSHEL	.09040	1.503
Air conditioning on bus:				
Air conditioning free	1,0	AVALFREE	.07131	1.112
AC at 20% extra fare	1,0	AVALPAY	-.17432	-2.207
Safety on board bus:				
Ride-generally smooth	1,0	GSBRAKE	.20788	2.963
Ride-very smooth	1,0	VSNBRAKE	.35232	4.904
Cleanliness of seats:				
Clean enough	1,0	CENOUGH	.13867	1.830
Very clean	1,0	VCLEAN	.20446	2.713
Ease of access to the bus:				
Wide entry/2 steps	1,0	WIDE2STP	.09589	1.499
Wide entry/no steps	1,0	WIDENSTP	-.10319	-1.372
Driver behavior:				
Driver friendly enough	1,0	FRIENDEN	.19798	2.572
Driver very friendly	1,0	VFRIEND	.42287	5.564
Information at the bus stop:				
Timetable only	1,0	TIMNOMAP	.29609	4.745
Timetable and map	1,0	TIMWMAP	.19720	3.021
Fequency of bus service:				
Frequency/every 60 mins	1,0	FREQ60	-.58595 -	6.902
Frequency/every 30 mins	1,0	FREQ30	-.12221	-1.640
Socioeconomic characteristics:				
Female	1,0	FEMALE	.09986	1.198
Personal income	$'000 s	PINCO	.00905	3.817
Age of passenger	years	AGES	.01379	5.787
Bus operator-specific dummy variables:				
Operator 1	1,0	Op1	.37358	1.671
Operator 2	1,0	Op2	.19642	654
Operator 3	1,0	Op3	-.94098	-5.497
Operator 4	1,0	Op4	-.17726	-1.080
Operator 5	1,0	Op5	-.12964	-.653
Operator 6	1,0	Op6	.97267	1.937
Operator 7	1,0	Op7	-.18127	-0.982
Operator 8	1,0	Op8	.35723	1.294

Table 3 Final user preference model (Continued)

Operator 9	1,0	Op9	-.26210	-1.215
Operator 10	1,0	Op10	-.56626	-1.845
Operator 11	1,0	Op11	-1.2555	-4.850
Operator 12	1,0	Op12	-.22189	-0.842
Operator 13	1,0	Op13	-.47366	-1.210
Operator 14	1,0	Op14	.01784	.072
Operator 15	1,0	Op15	.06911	.084
Operator 16	1,0	Op16	-.37973	-1.685
Operator 17	1,0	Op17	.06878	.292
Operator 18	1,0	Op18	-.36574	-0.825
Operator 19	1,0	Op19	1.1207	4.218
Operator 20	1,0	Op20	.10014	.488
Operator 21	1,0	Op21	.11275	.546
Operator 22	1,0	Op22	.32239	.781
Operator 23	1,0	Op23	-.53292	-1.845
Operator 24	1,0	Op24	.08878	.161
Log-likelihood	-2839.25			
Pseudo R^2 (adjusted)	0.324			

to be a major influence on service quality with both 'seats only' and 'bus shelter with seats' not being statistically significant relative to 'no shelter or seats'. If reproducible in further studies this has important policy implications as to priorities in service improvement. The availability of air conditioning is another interesting result. We find that 'air conditioning without a fare surcharge' is not statistically significant relative to no air conditioning. In contrast the provision of air conditioning with a 20% surcharge on existing fares is statistically significant with a negative sign suggesting that users would sooner not have air conditioning if it means paying higher fares.

On-board safety, defined by the smoothness of the ride is a statistically strong attribute. Relative to 'the ride is jerky with sudden braking occurring often', we find that 'the ride is generally smooth with rare sudden braking' and 'the ride is smooth with no sudden braking' are both very important positive attributes of service quality. This suggests both policy initiatives in driver skill as well as vehicle quality. Cleanliness of the bus is statistically significant when 'very clean' relative to 'not clean enough'. The non-statistical (1.830) significance of 'clean enough' suggests that we really have a dichotomy between very clean and not very clean. Ease of access to a bus, closely linked to the issue of accessible transport turns out to be not so important overall, presumably because the majority of users (including many aging users) are sufficiently healthy to not be concerned with the configuration of steps and entry widths. The attitude of the driver is a statistically strong influence on a user's perception of service quality. Indeed, relative to 'very unfriendly' we might expect a significant increase in the mean parameter estimate when we go from 'friendly enough' to 'very friendly'. This is the most non-linear effect on utility of all the attributes of service quality. The availability of information at the bus stop (timetable and map) is statistically important compared to

'no information', although surprisingly the key information item is a timetable, with a map being a liability (possibly because of experience with vandalism?).

Finally, bus frequency defined as 15, 30 and 60 minutes, was found to be significant when treated as a dummy variable distinguishing 60 minutes from 15 and 30 minutes. There is a strong negative sign for the 60 minute dummy variable, suggesting that a 60 minute service reduces relative utility significantly compared with a service frequency of every 15 or 30 minutes. Not statistically significant is the 30 minutes dummy variable, defined equal to one for frequencies equal to 30 minutes.

The socioeconomic characteristics sought from bus users (see Additional file 1: Table S3b) were limited to personal income, age, gender, occupation, and car availability. We found that individuals on higher incomes and of more years, were more likely to prefer the levels of service offered by the existing trip than by the alternative packages. What this suggests is that as individuals age and increase their income, they see existing service quality as increasingly satisfying their requirements for service quality. Alternatively, it is the younger users and those on lower incomes that see a greater need for improved service quality. Car availability was not statistically significant. Further details are given in Prioni and Hensher [25].

The customer service quality indicator (CSQI) and benchmarking

The CSQI for each operator is calculated by the application of the utility expression in Table 3 and the levels of each of the attributes associated with the current trip experience of each sampled passenger (as provided from Additional file 1: Table S3b of the survey). In this study we have estimated a single set of utility weights across the sample of 3,849 passengers using the services of 25 operators. We investigated possibilities of differences in weights between segments of operators (e.g., Sydney metropolitan vs. regional vs. country towns) and found no statistically significant differences. This is most encouraging, suggesting a similar pattern of preferences of passengers across all operating environments. This does not mean however that the levels of service offered on each service attribute are the same (indeed there is substantial variation in the mean and standard deviation of each attribute for each operator). Rather, what we are noting is that the marginal utility of each attribute (i.e., the mean parameter estimate of part-worth weight) is well represented by a single mean estimate across all operators.
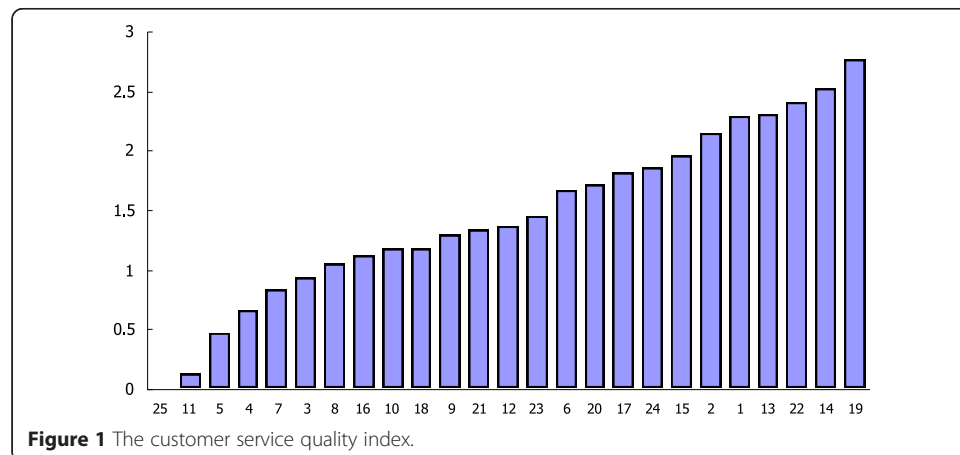
The aggregated CSQI developed for each operator is summarised in Table 4 and graphed in Figure 1 at its mean for each operator. We have normalised CSQI in Figure 1 to a base of zero for the operator with the lowest relative CSQI. The range is from 0 to 2.70.
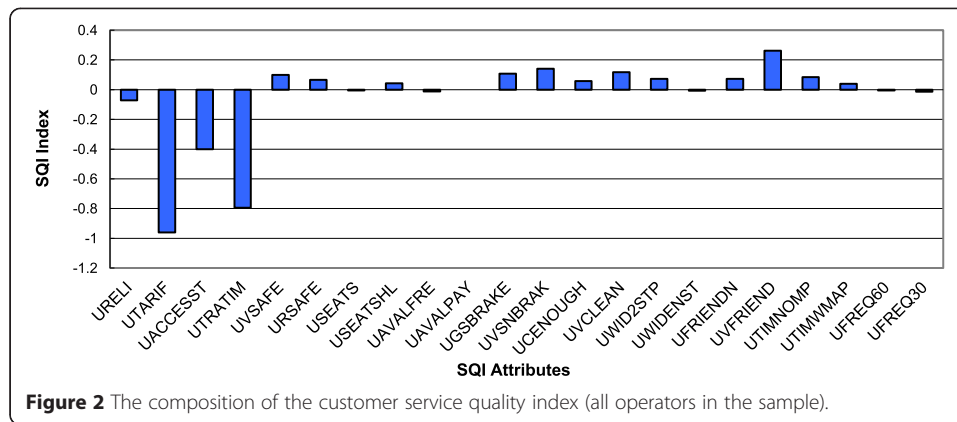
In developing the CSQI indicator, we have taken into account differences in the socio-economic composition of the travelling public (e.g., age, income, car availability) and location of each operator. The contribution of each service quality attribute across all 25 operators in summarised in Figure 2. The challenge for an operator is to compare themselves against best practice and to establish how best to improve overall service quality through implementing changes that reduce the magnitude of the attributes below the zero axis in Figure 2, and increase the magnitude of attributes above the zero axis. The parameters' estimates allow us to derive other interesting results. Figure 2 shows the contribution (in terms of utility), of each single quality attribute over the

Table 4 Summary statistics of customer service quality index

Operator	Mean	Standard deviation	Minimum	Maximum	Sample size
1	0.5311	0.788	-2.39	2.28	249
2	0.3900	0.894	-1.87	2.00	96
3	-.8178	1.248	-4388	1.92	508
4	-1.098	0.927	-5.58	0.58	374
5	-1.2840	1.406	-5.46	0.84	196
6	-.8377	0.383	-.525	0.80	24
7	-.9263	1.297	-6.74	1.82	412
8	-.7113	0.566	-2.12	0.44	150
9	-.4597	0.685	-2.55	1.06	173
10	-.5805	0.904	-3.06	0.67	64
11	-1.628	0.979	-4.55	0.55	90
12	-.3923	1.000	-3.80	1.40	100
13	0.5435	0.483	-.434	1.28	41
14	0.7636	0.940	-2.28	2.61	180
15	0.2079	0.637	-.638	0.692	9
16	-.6345	0.958	-4.00	1.03	159
17	-.0649	1.089	-2.86	2.09	190
18	-.5687	1.206	-3.24	1.04	27
19	1.0174	0.947	-.990	2.70	203
20	-.0444	0.639	-1.43	1.55	224
21	-.4212	0.852	-3.45	1.17	227
22	0.6466	0.643	-.600	2.01	46
23	-.3076	1.034	-4.28	.808	65
24	.1051	1.156	-2.17	1.42	22
25	-1.7579	.875	-3.01	-.096	20
All	-.4067	1.224	-6.74	2.70	3849

Note: the 'All' values are based on summing across the entire data set of 3,849 bus users and not based on a simple average of the 25 values reported for each operator.

Figure 1 The customer service quality index.

Figure 2 The composition of the customer service quality index (all operators in the sample).

entire sample (see Table 3 for the complete list of attributes and acronyms). Tariff (UTARIF), travel time (UTRATIM) and access time (UACCESST) have the highest impact on service quality On the positive side of CSQI, the major influence is given by the friendliness of the driver (UVFRIEND) and the smoothness of the ride (UVSNBRAK).

Crucially, in assessing the performance of each operator on CSQI, we must identify those attributes that the contracted operator has no control over and ensure that any benchmarking of their performance separates out these attributes and only benchmarks on the basis of attributes the operator has control over. The excluded attributes however still reveal very useful information since it is an indication of the overall state of consumer satisfaction regardless of who has control of making changes to improve customer satisfaction. In a real sense then, we have identified the full story in respect of each contracted regime regardless of who has control to effect changes in each of the influencing attributes. This is what matters to the customer, and hence the full suite of measures underlying CSQI become the relevant set to the ultimate stakeholder - the end users or passengers.

When we presented the findings to the operators, they found them not only illuminating but also of real practical value in guiding them on where to focus service improvements in order to obtain higher customer satisfaction and an improved benchmarked CSQI. An area that was acted on immediately by many operators was increased and different training of drivers to ensure that the way they respond and support passengers was improved. It was also found that where a driver remained on the same route, they got to know their passengers much better than a roster that involved moving between many routes. By supporting retention of drivers to a few routes, customer relationships improved significantly (it was even suggested that we cannot afford to rude to a passenger because they will see us the next day).

In Figure 3 we illustrate the types of useful information that is in the CSQI for each operator. Given knowledge of which attributes are under the control of the operator they can identify which attributes they might improve on in order to improve on their overall CSQI. For example, it is worth noting that the first 4 attributes from the left are almost certainly not under the control of the operator in the Sydney context (i.e., travel time reliability, fare, access time and travel time), although one might suggest that
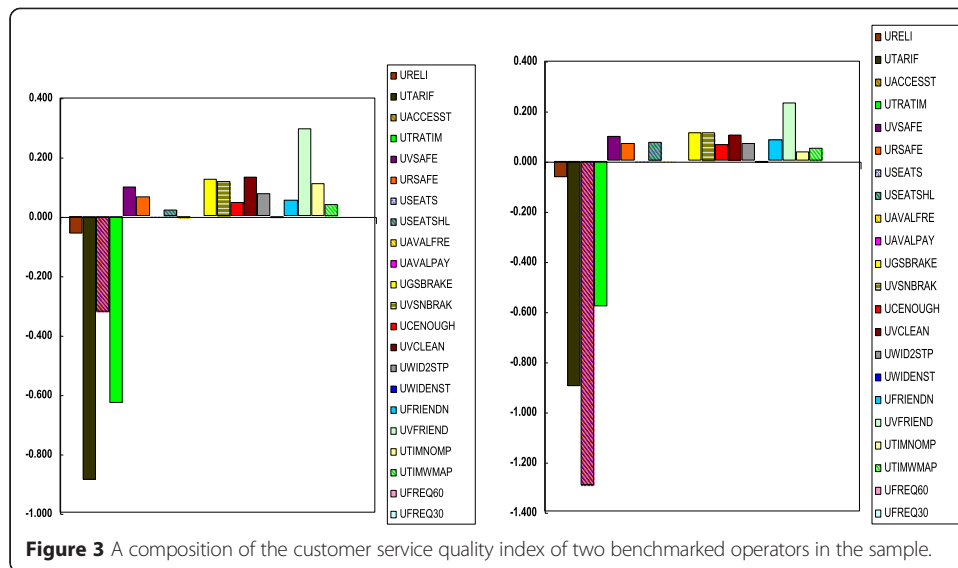
Figure 3 A composition of the customer service quality index of two benchmarked operators in the sample.

access time could be changed through greater spatial connectivity which the operator has some control of above contract specified minimum service levels. Hence the set of influences below the zero line are essentially out of operator control, which means that they need to focus on improving the attribute levels above the line. The very friendly attitude of the driver is clearly a strong contribution to positive service quality for both operators, and is possibly something that is relatively easy to enhance and hence improve the overall CSQI.

To conclude this section we use a case study undertaken in Singapore where the method was also applied. The interest was in benchmarking the performance of an operator over time. Taking two years as shown in Table 5 (including the estimated multinomial logit model), the calculation of the CSQI using the average attribute levels for each year in Table 5 from the sample of users of the specific operator's services, is:

$$\text{CSQI}_{\text{year 1}} = \text{Utility}_{\text{year 1}} = 1.14 - 0.04*95 - 0.2*22 - 0.41*8 + 0.65*(\text{-}1)$$
$$+ 0.51*1 + 0.31*1 = \text{-}10.17$$
$$\text{CSQI}_{\text{year 2}} = \text{Utility}_{\text{year 2}} = 1.14 - 0.04*120 - 0.2*22 - 0.41*6 + 0.65*1$$
$$+ 0.51*1 + 0.54*1 + 0.31*1 = \text{-}8.51$$

Negative utilities do not matter (we could have normalized to be positive as above); what matters is the change (+ or −) from the baseline. The evidence indicates that service quality improved between Year 1 and Year 2.

Conclusions

Putting the customer at the centre of policy, planning and delivery decisions requires a measure of customer satisfaction that is robust, capable of benchmarking customer satisfaction over time to inform policy and planning of services, and to inform how the separate elements of service quality contribute to service performance with a view to informing contract management, recognising the need to distinguish between attributes of service that are reasonably under the control of the operator, the regulator and the market.

This paper has developed an improved way of recognising the packaging of service quality attributes in the delivery of bus services under government contracts. In moving away from univariate measures of customer satisfaction associated with singularly

Table 5 Comparison of overall service performance between two years for a Singapore operator

Attributes	Coefficient	Attributes	Coefficient
Constant	1.14* (0.22)	Seated part of the way	−0.11 (0.11)
Fares	−0.04* (0.0066)	Very smooth journey	0.51* (0.11)
Travel Times	−0.20* (0.027)	Slightly jerky journey	0.42* (0.11)
Waiting Time	−0.41* (0.076)	Information (Reliable)	0.16 (0.13)
Very Clean	0.54* (0.11)	Information (Unreliable)	−0.10 (0.12)
Moderately Clean	−0.19 (0.12)	Facilities	−0.11 (0.082)
Temperature just nice	0.31* (0.11)	Range of Waiting Time	−0.035 (030)
Seated all the way	0.65* (0.10)	* denotes statistical significants at 5% level	
YEAR 1		**YEAR 2**	
Components	**Data**	**Components**	**Data**
Fares	95 cents	Fares	120 cents
Travel Time	22 mins	Travel Time	22 mins
Waiting Time	8 mins	Waiting Time	6 mins
Seat Availability	Standing all the way	Seat Availability	Seated all the way
Smoothness of Ride	Very smooth journey	Smoothness of Ride	Very smooth journey
Cleanliness	Moderately Clean	Cleanliness	Very Clean
Temperature	Just Nice	Temperature	Just Nice

defined attributes to the mix of attributes offered in a bus service, we use a discrete choice multinomial logit model to identify the role that each attribute in a package ('an alternative') plays in defining the level of utility (or satisfaction) applicable to each bus travelling member of the population.

Importantly the stated preference method has as its sole purpose an enriched strategy to ensure a richer understanding of preferences for attributes describing service in a situation where the variability is actual service levels may not be rich enough to enable a fully revelation of customer preferences for each attribute. Then implementing an estimated model on actual experiences (revealed preference data) enables us to identify customer satisfaction overall from the package of experienced attribute levels, named the customer service quality index.

This index when aggregated across a sample of users of an operator's services enables us to obtain an operator specific CSQI, which can be benchmarked against other operators, distinguishing between those attributes that have levels under the control of the operator and those that are controlled by the regulator or the market at large. All attributes, no matter who has 'responsibility for their level', clearly matter to varying degrees to the end user and hence must all be taken on board in changes designed to create additional value in the bus use experience, and consequent justification through value for money of the taxpayers outlay to support existing and improved bus services under government contracts.

In this paper we have quantified CSQI based on customer feedback from the performance of incumbent operators. This can be used to monitor the performance of incumbents under contract with the proviso that the contributing influences to the overall CSQI are distinguished in respect of who has control of and hence is responsible for each underlying source of relative satisfaction. However, this approach is also valuable in a competitive tendering or negotiated contract setting in that it can be used

to set targets or standards that are aligned with what is already achieved (or better) under current contracts that are being replaced.

Endnotes

[a]Eboli and Mazzulla [3] is a very useful review of the literature on methods used to study public transport service quality.

[b]The MNL and more advanced methods are discussed in detail in Hensher et al. [12].

[c]School children were excluded from the sample, as they are captive users and might have a biased perception towards the attributes.

Appendix

Notation

URELI Late minutes
UTARIF Bus fare
UACCESST Access time
UTRATIM Travel time
UVSAFE Very safe
URSAFE Reasonably safe
USEATS Seats only at bus stop
USEATSHEL Seats plus shelter at stop
UAVALFREE Free Air conditioning
UAVALPAY Air conditioning at 20% extra fare
UGSBRAKE Smooth ride
UVSNBRAKE Ride very smooth
UCENOUGH Clean enough
UVCLEAN Very clean
UWIDE2STP Wide entry and 2 steps
UWIDENSTP Wide entry no steps
UFRIENDN Friendly drivers
UVFRIEND Drivers very friendly
UTIMWMAP timetable and map
UTIMNOMAP Timetable, no map
UFREQ60 Frequency 60 minutes
UFREQ30 Frequency 30 minutes

Additional file

Additional file 1: Table S3a. A Typical Stated Preference Exercise. **Table S3b.** Revealed Preference data collected.

books. David has advised numerous government and industry agencies in many countries (notably Australia, New Zealand, UK, USA and The Netherlands), with a recent appointment to Infrastructure Australia's reference panel on public transport, and is called upon regularly by the media for commentary.

References

1. Cirillo C, Eboli L, Mazzulla G (2011) On the asymmetric user perception of transit service quality. Int J Sustainable Transportation 5(4):216–232
2. dell'Olio L, Ibeas A, Cecìn P (2010) Modelling user perception of bus transit quality. Transport Policy 17:388–397
3. Eboli L, Mazzulla G (2010) How to capture the passengers' point of view on a transit service through rating and choice options. Transport Rev 30(4):435–450
4. Bolton RN, Drew JH (1991) A multistage model of customers' assessment of service quality and value. J Consum Res 17:375–84
5. Boulding W, Kalra A, Staelin R, Zeithaml VA (1993) A dynamic process model of service quality: from expectations to behavioral intentions. J Mark Res 30:7–27
6. Hensher DA, Prioni P (2002) A Service Quality Index for Area-wide Contract Performance Assessment. J Transport Econ Pol 36:93–113
7. Hensher DA, Stopher PR, Bullock P (2003) Benchmarking and service quality at a market segment level. Transport Res Part A 37:499–517
8. Buttle F (1996) SERVQUAL: review, critique, research agenda. Eur J Mark 30(1):8–31
9. Parasuraman A, Zeithaml V, Berry LL (1985) A conceptual model of service quality and its implications for future research. J Mark 49:41–50
10. Andersson TD (1992) Another model of service quality: a model of causes and effects of service quality tested on a case within the restaurant industry. In: Kunst P, Lemmink J (eds) Quality Management in Service. van Gorcum, The Netherlands, pp 41–58
11. Richard MD, Allaway AW (1993) Service quality attributes and choice behavior. J Serv Mark 7(1):59–68
12. Hensher DA, Rose JM, Greene WH (2005) Applied Choice Analysis: A Primer, Cambridge University Press. Second edition published in 2015, Cambridge
13. Hensher DA (2014) The relationship between bus contract costs, user perceived service quality and performance assessment, (presented at Thredbo 12, Durban, South Africa, September 2011). Int J Sustainable Transportation Spec Issue 8(1):5–27
14. Hensher DA (1992) Performance evaluation and passenger transit: what are the relevant measures? The Proceedings of the Second International Conference on Privatisation and Deregulation of Passenger Transport, Finland, pp 61–72
15. Fielding GJ, Babitsky TJ, Brenner ME (1985) Performance evaluation for bus transit, Transportation Research, 19 (1), 73– 82.
16. Kittleson and Associates (1996) Development of Transit Capacity and Quality of Service Principles, Practice and Procedures, *Transit Cooperative Research Program Project A-15*. Transportation Research Board, Washington D.C
17. Swanson J, Ampt L, Jones P (1997) Measuring bus passenger preferences, Traffic Engineering and Control, June, 330–36.
18. Cunningham LF, Young C, Lee M (1997) Developing customer-based measures of overall transportation service quality in Colorado: Quantitative and qualitative approaches. J Public Transportation 1(4):1–22
19. dell'Olio L, Ibeas A, Cecìn P (2010a) The quality of service desired by public transport users, Transport Policy, doi:10.1016/ j.tranpol.2010.08.005.
20. Eboli L, Mazzulla G (2008) An SP experiment for measuring service quality in public transport. Transportation Plann Technol 31(5):509–523
21. Eboli L, Mazzulla G (2008) Willingness-to-pay of public transport users for improvement in service quality. Eur Transport 38:107–118
22. Marcucci E, Gatta V (2007) Quality and public transport service contracts. Eur Transport 36:92–106
23. Louviere JJ, Hensher DA, Swait J (2000) Stated Choice Methods: Analysis and Applications in Marketing, *Transportation and Environmental Valuation*. Cambridge University Press, Cambridge
24. Train K (2009) Discrete Choice Methods with Simulation. Cambridge University Press, Cambridge
25. Prioni P, Hensher DA (2000) Measuring service quality in scheduled bus services. J Public Transport 3(2):51–74