Transportation Research Part A 86 (2016) 108–123

Contents lists available at ScienceDirect

Transportation Research Part A

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

The best of times and the worst of times: A new best–worst measure of attitudes toward public transport experiences

Matthew J. Beck [a,*], John M. Rose [b]

[a] Institute of Transport and Logistics Studies, The University of Sydney, 378 Abercrombie St, Darlington, NSW 2006, Australia
[b] Institute for Choice, University of South Australia, Australia

ARTICLE INFO

Article history:
Received 16 January 2015
Received in revised form 15 December 2015
Accepted 8 February 2016

Keywords:
Best worst scaling
Service quality
Attitudes
Public transport
Satisfaction
Importance

ABSTRACT

Attitudes play an important role in determining individual transit behaviour and the measurement of attitudes is relied on by public transit authorities' world over. Given their role in behaviour and policy making, the accurate measurement of attitudes is of critical importance. Traditional satisfaction scales are prone to bias and on their own they are only a partial measure of attitudes. Given that satisfaction scales have been used to assist with large scale transport infrastructure investment decisions, to aid policy makers examining reactions to alternative policy changes and reform, and to measure the success of new initiatives, deriving robust satisfaction scales should be of critical importance. This paper introduces a dual version of best–worst scaling as an alternative measure of satisfaction. Best–worst scaling is free of the biases inherent in traditional response scales and is ideal for handling the comparative evaluation of large amount of attributes, particularly those which are inherently qualitative. The paper makes a further innovative contribution by proposing a model structure for the joint estimation of satisfaction and importance. Our model shows a better delineation between the attributes used to measure attitudes towards bus use and a more detailed understanding of the relationship between importance and satisfaction; enabling transport operators to better understand what counts most and assess their performance.

1. Introduction

The role of attitudinal data in the transportation literature has long been established (Paine et al., 1969; Recker and Stevens, 1976) where early research found that attitudes may be better predictor of modal choice than more objective measures (Gilbert and Foerster, 1977). Since then, research has focused on better understanding the role attitudes play in terms of public transport use, particularly in terms of the motivations of users and non-users, with particular reference to promoting ridership (Beirao and Cabral, 2007) and gaining improved societal outcomes with respect to social mobility (Bouf and Hensher, 2007). Thompson and Schofield (2007) explore the way in which attitudes towards public transport performance can influence satisfaction with the destination being travelled to. Chou et al. (2014) use structural equation modelling to identify which attitudes significantly affect satisfaction with high speed rail services in Taiwan. Zhang et al. (2014) model a subjective evaluation of bus comfort as a function of objective characteristics of the bus journey such as noise and vibration. A classification and regression tree approach (CART) has been used to identify the characteristics influencing overall

* Corresponding author. Tel.: +61 2 9114 1834.
E-mail addresses: matthew.beck@sydney.edu.au (M.J. Beck), john.rose@unisa.edu.au (J.M. Rose).

http://dx.doi.org/10.1016/j.tra.2016.02.002
0965-8564/

service quality (de Ona et al., 2014, 2015). The attitudes towards desired level of service, as opposed to the actual level of service experienced, have also been explored (Dell'Olio et al., 2011).

While the academic literature has focused on the role of attitudes, their measurement plays are a critical role in the ongoing evaluation of existing public transport links. The Victorian Department of Transport in Australia have found that research into the public's attitudes towards transport plays a "critical role" in informing and guiding policy, that changes to attitudes led to a significant increase in public transport patronage between the years 2006 and 2008 (Gaymer, 2010). In the state of New South Wales the Transport Customer Satisfaction Index Survey is designed to gather information about traveller satisfaction with train, bus and ferry services so as to gauge the strength of public opinion on service attributes such as accessibility, timeliness, cleanliness, information, comfort, ticketing, safety and convenience. Similarly, the UK's Department for Transport and New York City's Metropolitan Transit Authority both make extensive use of customer satisfaction surveys to identify key areas for improvement and management attention, and to gauge reaction to new services and initiatives.

Given the importance that attitudes play in determining transit behaviour and evaluating the performance of public transport networks, it is important that attitudinal research not only explore the role that they play, but also examine alternative approaches to survey response mechanisms that may result in more robust, as well as managerially more useful attitudinal data. A commonality among the traditional studies of attitudes, as well as the data used in practice, is the method via which attitudinal information is extracted, that being a psychometric 5-point Likert scale, or a close variant thereof, where respondents are asked to provide their level of agreement/disagreement or satisfaction/dissatisfaction across a range of attitudes, perceptions or experiences. While the use of these types of questions are widespread and heavily relied upon by many public administrators, policy designers and service providers in the formation of important policy and business decisions, these unit-scale style questions are subject to several fundamental criticisms.

The respondent's involvement with such survey questions, how engaged they are when completing the survey, situational factors in the environment surrounding them as complete the survey, and how accessible or easily recalled the attitudes are for respondents can all influence the responses provided (Mowen, 1993). The evaluation of such a ratings scale itself is subjective and how a respondent evaluates their position on that scale differs across respondents – for example, what is enough to make you very dissatisfied maybe more or less than what it takes for someone else to be very dissatisfied. Such variation in response styles has been shown to significantly affect the means and variance of the estimates obtained from these types of surveys (Craig and Douglas, 2000; Steenkamp and Baumgartner, 1998). These types of questions allow respondents to enact decision shortcuts or simplification strategies such as saying everything is "good" or everything is "bad" (i.e., not truly consider the spectrum of possible results) without penalty. This type of response style makes it extremely difficult to determine the most important issue or understand the priority of different issues. Additionally, research has also shown that another three potential response biases can occur with such rating scales; social desirability bias, acquiescence bias, and extreme response bias (Paulhus, 1991). It is not inconceivable that attitudes towards public transport may be highly prone to all three of these.

Given these known issues with scale response questions, alternative methodologies that allow for a potentially less biased examination of service quality have been proposed and tested within the literature. For example, Hensher et al. (2003) proposed a service quality index that uses stated choice methods. Passengers were asked to choose their most preferred "bus package" from a number of alternative packages of service levels based on thirteen attributes of a typical bus journey. The resultant multinomial logit models established the relative weights attached to the statistically significant attributes, representing the contribution of each service attribute to the calculation of an overall service quality index. This process has been repeated by others in different contexts (Eboli and Mazzulla, 2008; Roman et al., 2014) and has been used to show that consumer preferences over bus service attributes are non-linear (Stathopoulos and Marcucci, 2014); a result that would be typically masked in traditional response scale based studies. While this study shows merit, it is difficult to incorporate a large number of service attributes into the design of the experiment, and explaining the qualitative or "soft" factors in a meaningful way in a stated choice experiment can be problematic.

Recognising the fundamental issues with rating-scales more directly, an alternative approach for the collection of attitudinal data can be recommended. First proposed by Finn and Louviere (1992), the elicitation method known as best–worst scaling (BWS) requires respondents to choose the "best" and the "worst" option from a limited set of statements; which themselves are extracted from a larger "master set". Each respondent is provided with a number of these choice sets, with statements shown in different combinations. This type of approach is becoming increasingly popular in marketing, healthcare and welfare analysis because of the distinct advantages this method offers over the traditional approach of rating scales. With transportation, the role of BWS as a more precise measure of attitudes is slowly gaining attention: it has been used to examine improvements to New York City Transit's subway stations (Spitz et al., 2007) and the Syracuse Metropolitan Transportation Council used BWS to evaluate service improvement priorities (SMTC, 2011). The method has been adopted by leading transportation consultants (RSG, 2013) and is viewed by the Transportation Research Board as an innovative tool for understanding how individuals make decisions (TCRP, 2008).

In identifying the benefits of BWS, Cohen and Markowitz (2002) state that there is only one way to choose the most important item and, as such, respondents cannot consistently use the middle, the end points, or one end of the importance scale, forcing discrimination among the items. The method offers the opportunity to evaluate attributes relative to each other (Lee et al., 2008), allows the measurement of both attitudes towards an object and towards a behaviour (Flynn et al., 2008), and produces reliable and interpretable estimates of the relative impact of attribute levels (Marley et al., 2008). Additionally, for certain types of BWS methods (termed Case 1 and Case 2 in the BWS literature) all items are measured on a common

scale (Auger et al., 2007), with results that closely approximate the true scale values obtained from logit analysis (Marley and Flynn, 2005).[1] As a result, scalar differences generated by the way people interpret rating scales can be overcome (Cohen and Neira, 2003); results are more sensible and discriminating than rating scales (Lee et al., 2007); better sample discrimination is possible (Mielby et al., 2012); more information about an individual's ranking can be gathered, replies are more unambiguous as people are generally clearer about the extreme options, and the questionnaire task appears easier for the respondent relative to other methods (Marley and Louviere, 2005; Marley, 2010).[2] Finally, and particularly relevant to this study, BWS is conceptually better at determining the relative impact of a large number of attributes, particularly qualitative effects. Readers who are interested in a fuller discussion of the known properties of best worst methods are referred to Marley and Flynn (2015) and Flynn and Marley (2014).

This paper seeks to contribute to the wider study of attitudes. While it is widely accepted that attitudes are a function of both belief and importance (see e.g., Fishbein and Ajzen, 1975), the typical satisfaction survey used in the transportation examples discussed above, as well as studies in marketing, health, environment, and psychology for example, collect only one half of this equation (in this case belief). Unless the importance of each attribute, feature, statement or construct is also collected you do not truly measure attitude. The structure proposed in this paper allows for the parsimonious collection of both satisfaction and importance and thus allows for the joint measurement of both constructs; a task that is much more difficult, if not impossible, using other methods.

The objectives of this paper are twofold. Firstly, we introduce BWS as an alternative method for measuring attitudes and subsequently the performance of public transit systems; and to propose a methodology for the joint specification of satisfaction and importance. The remainder of the paper is structured as follows. In Section 2, we discuss the sample used for the current study as well as the survey tasks respondents were asked to complete. Also discussed in Section 2 are the results of a traditional ratings task examining satisfaction and importance for 25 items related to bus usage. Next, Section 3 outlines the modelling methodology used for the new service attitude index task, after which the results are presented in Section 4. Section 5 concludes the paper with some final observations.

2. Data

2.1. Survey description

Respondents completing the survey were first asked a series of screening questions related to their current level of bus usage as well as questions about their most recent experience using a bus (see Table 1). Next, respondents were presented with a series of traditional ratings based tasks designed to elicit their perceived degree of importance and satisfaction with respect to their experience of bus travel in general. A list of factors that contribute to the overall experience of bus transit was developed via a review of the extant literature. The list comprised of factors relevant to the following stages of a journey: before the trip was undertaken (e.g., timetable information); while waiting for the bus (e.g., quality of the stop); accessing the bus (e.g., ease of boarding); travelling on the bus (e.g., comfort of the seat); and at the destination (e.g., ease of transfer). This list was refined to 25 key statements via in-depth interviews and pilot studies with both bus users and bus operators. The authors strongly recommend that for any study of attitudes to be truly complete, exhaustive qualitative research is conducted in order to really uncover all the relevant attitudes.

After completing the ratings tasks, respondents completed 10 best worst case 1 tasks, consisting of four statements drawn from the 25 they previously provided importance and satisfaction ratings. Within the best–worst case 1 method, the number of attributes shown in any given task is determined such that there are enough attributes for respondents to trade between, but not so many that the respondent finds the task too difficult (or to trivial). In this experiment, pilot surveys revealed that four attributes was adequate to allow respondents to trade between the best and the worst with sufficient accuracy, while still managing to present respondents with a small enough number of tasks overall to keep them engaged. As with other choice experiments, the number of tasks required to show a single respondent all combinations of attributes is unfeasibly large. To reduce the choice tasks required for each respondent such that over the sample every pairwise combination of service attributes is shown, a balanced incomplete block design (BIBD) was used.[3] A total of 50 choice sets were constructed, with this master set being blocked into five blocks of 10 choice tasks each. To block the design, an algorithm was written to enforce the rule that each statement had to appear at least once in each block, such that every respondent had the opportunity to rate that service attribute as the best or the worst. An example screen capture of the task is provided in Fig. 1.

[1] For many sets of BWS data (Cases 1, 2 or 3), there tend to be (highly) linear relations between best minus worst scores and the corresponding parameter estimates for logit-type models. However, the slope of that linear relation is often not equal to one and differs from study to study; hence one cannot immediately obtain the "true" parameter estimates by simply taking the best minus worst scores.

[2] Due to the ability of BWS to reduce scalar inequalities the method also facilitates better cross-regional comparisons (see Auger et al. (2007) for a more detailed discussion). This property is potentially appealing to those in transport who seek to examine internationally diverse attitudes and behaviours, which may be particularly relevant for those studying attitudes towards public transport.

[3] An incomplete block design is a type of design where the number of feasible alternatives exceeds the number of alternatives that can be presented in any one task (in the experimental design literature dealing with incomplete block designs, each choice task is the equivalent of a block). A BIBD is an incomplete block design in which each paired combination of alternatives appears an equal number of times over the design (choice tasks or blocks). There exist a finite known number of BIBDs. The BIBDs for this study were obtained from a research version of the software Ngene.

Table 1
Bus usage descriptive statistics.

How long ago did you last caught a bus?	
Last used a bus today	26.37%
Last used a bus yesterday	34.33%
Last used a bus a week ago	27.86%
Last used a bus a fortnight ago	5.47%
Last used a bus a month ago	5.97%
How often do you use a bus?	
Use a bus 5–7 times a week	31.34%
Use a bus 3 or 4 times a week	29.35%
Use a bus 1 or 2 times a week	27.86%
Use a bus once a fortnight	11.44%
What was the main purpose of your last bus trip?	
Commuting	46.77%
Shopping	18.91%
Business	11.44%
Education	6.47%
Visiting friends or family at their home	4.98%
Going to a public event (e.g., concert, watch a sporting match)	3.98%
Day trip	2.99%
Going to a social event in a public area (e.g., a movie, dinner or drinks)	2.49%
Other	1.99%

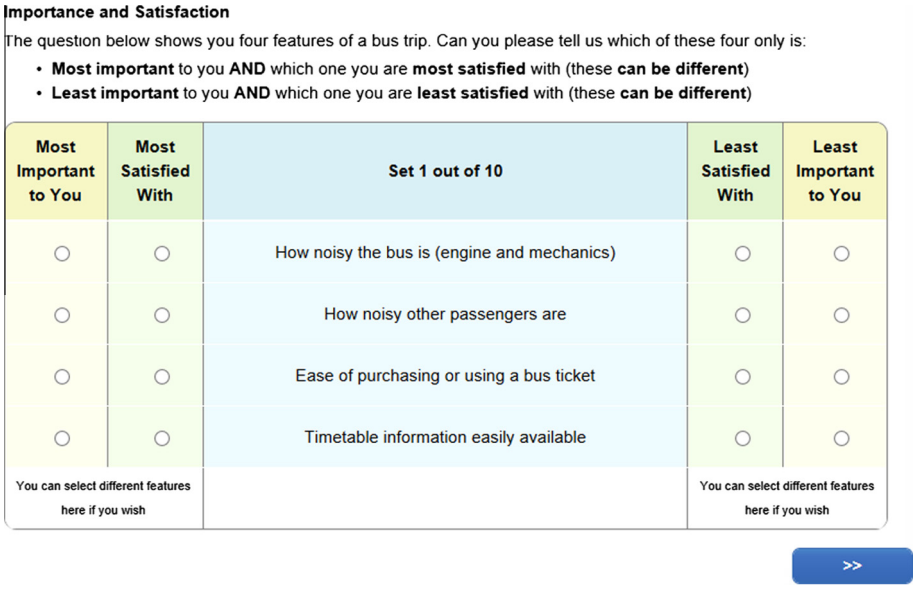

Fig. 1. Best–worst measurement of attitude.

The questionnaire concluded with a series of questions about the socio-demographic characteristics of the respondent (see Table 2). The median time respondents took to complete the survey was 9.35 min, with a median time of 24 s to complete each response task (i.e., the joint satisfaction and importance best worst questions).

2.2. Sample description

Eligible respondents were limited to persons aged 18 years or older who, at the time of the survey, had access to a bus service, whether that bus service was used or not. Access to a bus service was not limited to trips originating from the persons home. Data were captured from 252 respondents, 201 of whom had reported using a bus at least once in the fortnight prior to undertaking the survey. Sampled respondents were recruited using an internet marketing research panel (www.gmi-mr.com), drawn from the population of Sydney Australia. For the current paper, we make use of the data collected from the 201 bus users only.

Table 1 presents the summary statistics from this sample in terms of their reported bus usage. Sixty-one percent of respondents reported using a bus either the day of, or the day before completing the survey, with an additional 27.86 percent

Table 2
Sample descriptive statistics.

Age (years)	38
Income ($'000)	80
Gender (female)	43.79%
Full time employment	55.22%
Part time employment	14.43%
Casual employment	5.97%
Not employed	24.38%
Full time student	18.91%
Part time student	8.46%

of respondents stating that they had used a bus in the previous week. Thirty-one percent of respondents stated that they use a bus at least five to seven times a week, whilst a further 29.35 percent reported using a bus at least three to four times a week on average. An additional 27.86 percent of respondents claim to use a bus once or twice a week, with the remaining 11.44 percent of respondents reportedly using a bus at least once a fortnight. With regards to the last bus trip, the largest number of respondents reported the primary purpose of the trip as travelling to or from work, followed by shopping and then business travel.

Table 2 presents a summary of the descriptive statistics for the 201 respondents who reported using a bus in the previous fortnight. The median age of the sample was 38 years, compared to 36 for the population of Sydney (ABS, 2011). Of the sample, 44.79 percent were female (compared to 50.77 percent for population of Sydney) with a median income of $80,000 per year ($75,244 for population of Sydney). Fifty-five percent of the sample reported being in full time employment, whilst 27.37 percent are students, either full time or part time.

2.3. Rating scale results

For the ratings tasks, respondents were asked to rate on a five point Likert scale their degree of satisfaction as well as how important each of the 25 items is to them. The importance and satisfaction ratings scales were collected for the purpose of comparing the results with the best worst experiment. Table 3 presents the mean and standard deviation for each of the 25 ratings tasks, as well as correlations between the corresponding satisfaction and importance items. Five of the 25 items were found to have a median of five out of a maximum five for importance ratings, with the remaining 20 items each had a median value of four. All 25 items were observed to have a median rating of four out of five for the satisfaction questions. Examining the table further, for all 25 items, the mean importance value is greater than the mean satisfaction rating. The largest correlation observed in the data was 0.54 (for information available on the website) with the smallest correlation observed to be 0.1 (for cost of bus trip). The average correlation over all items was 0.34. To test whether the mean values are different for both ratings tasks, a series of paired sample *t*-tests were conducted on the data. As shown in the table, with the exceptions of ease of boarding, ride comfort, cleanliness of the bus stop at the end of the journey, and whether shelter is provided or not at the end of the journey, the mean importance ratings are statistically different (and larger) than the mean satisfaction rating for the same items. This suggests that respondents rate the majority of items more highly on importance than they do in terms of satisfaction.

In order to determine whether differences in the mean ratings for both the importance and satisfaction scales exist, an ANOVA test was conducted. Whilst not reported, statistically significant differences in means for both importance and satisfaction tasks were found. A Tukey's Honestly Significant Difference (HSD) Test (Tukey, 1949) was then performed *post hoc* to determine which items are statistically different. The results of this test, in the form of *p*-values, are shown in Table 4. To read the table, each column and row represents a separate item (numbered as per Table 3), with the cells of the table representing the *p*-value for pairwise differences between the differently numbered items. Results of the test related to the importance ratings tasks are presented below the leading diagonal (shown in light blue), whilst above the leading diagonal are the results associated with the satisfaction items (in light green). For example, to determine if the mean for item 2 (location of the bus stop at the start of the journey) is different to the mean importance rating for item 16 (comfort of seat), one simply reads down the second column until they locate the cell associated with the 16th row. In this instance, the *p*-value is 0.007 suggesting that the means of the two items are statistically different in terms of the importance rating (examining Table 3, it is clear that the bus stop location at the start of the journey (mean = 4.35) is statistically greater in terms of importance than how comfortable the seat is (mean = 3.89). To compare whether the means for the same two items are different for the satisfaction ratings tasks, the half of the table above the leading diagonal is employed. Reading across row 2 and down column 16, the *p*-value is 0.412, suggesting that one cannot reject the null hypothesis for the test, and hence results in the conclusion that the means for the two items are not statistically significant when looking at the satisfaction ratings.

Of note, only 15 or five percent of the possible 300 pairwise differences are statistically significant for the importance ratings, and 13 or 4.33 percent of the possible pairwise differences are statistically significant for the satisfaction ratings. Whilst there exist several possible reasons for this outcome, the two most likely are that either respondents did not properly engage with the task tending to select similar values for each item, or alternatively respondents viewed all items as equally

Table 3
Importance and satisfaction ratings task results.

Statement	Importance			Satisfaction				
	Mean	Median	Std dev.	Mean	Median	Std dev.	*Corr(Imp,Sat)*	*t*-test of diff.
1. Information available on website	4.21	5	1.18	3.85	4	1.10	0.54	4.757
2. Location of bus stop (at start of journey)	4.35	5	1.10	3.99	4	0.98	0.47	4.713
3. Cleanliness at bus stop (at start of journey)	4.01	4	1.09	3.80	4	1.03	0.36	2.537
4. Signage at bus stop (at start of journey)	4.09	4	1.13	3.80	4	1.07	0.38	3.411
5. Shelter provided at bus stop (at start of journey)	4.15	4	1.10	3.71	4	1.13	0.32	4.832
6. Frequency of bus service	4.31	5	1.11	3.52	4	1.20	0.31	8.335
7. Bus being on time (keeping to timetable)	4.37	5	1.07	3.52	4	1.23	0.23	8.429
8. Ease of boarding the bus	4.03	4	1.05	3.97	4	1.05	0.45	**0.771**
9. Ease of payment	4.14	4	1.11	3.76	4	1.12	0.31	4.092
10. Cost of bus trip	4.19	5	1.16	3.47	4	1.24	0.10	2.522
11. Cleanliness of the bus	4.14	4	1.07	3.90	4	1.04	0.35	2.855
12. Friendliness of staff	4.11	4	1.02	3.71	4	1.07	0.35	4.866
13. Knowledgeability of staff	4.08	4	1.06	3.60	4	1.11	0.24	5.165
14. Level of crowding	4.01	4	1.11	3.73	4	1.11	0.35	3.218
15. Ability to get a seat	4.00	4	1.07	3.80	4	1.06	0.40	2.544
16. Comfort of seat	3.89	4	1.05	3.68	4	1.12	0.20	2.213
17. Leg room on bus	3.91	4	1.06	3.64	4	1.08	0.33	3.082
18. Comfort of ride	3.87	4	1.12	3.70	4	1.04	0.35	**1.951**
19. Noise on bus	3.95	4	1.10	3.53	4	1.12	0.30	4.456
20. Travel time	4.05	4	1.09	3.72	4	1.17	0.36	3.675
21. Temperature	3.97	4	1.12	3.75	4	1.11	0.32	2.334
22. Location of bus stop (at end of journey)	4.19	4	1.12	3.92	4	1.02	0.46	3.397
23. Cleanliness at bus stop (at end of journey)	3.95	4	1.11	3.84	4	0.98	0.33	**1.281**
24. Ease of connection to next service (at end of journey)	4.14	4	1.09	3.91	4	1.16	0.39	2.663
25. Shelter provided at bus stop (at end of journey)	3.94	4	1.18	3.77	4	1.11	0.32	**1.792**

Bolded font represent insignificant results at the 5% level.

important and are equally satisfied with the Sydney bus system. Without further evidence, we do not offer an opinion as to which is the most likely of the two possibilities.

To further investigate the similarity in responses we conducted confirmatory factor analysis, which revealed that for importance, all of the 25 service attributes loaded onto a single (unrotated) factor which explained 71.435 percent of the overall variance (min loading 0.773), with no rotated factor possible to compute. Cronbach's alpha coefficient was used to assess reliability for the whole importance scale. This revealed a value of 0.983, which indicated satisfactory reliability. We did manage to find two factors for the satisfaction items, the first of which explained 61.82 percent of the variance, and the second 4.11 percent. All items except item 10 (Cost of bus trip) were found to load onto the first factor, whilst all items save item 23 (Cleanliness at bus stop (at end of journey)) were found to load onto the second factor. As with the importance scale, the reliability of the scale was tested by computing a Cronbach alpha coefficient. For the satisfaction questions, a Cronbach alpha coefficient of 0.974 was computed, again indicating a high degree of reliability amongst the items. This conforms to the ANOVA results previously discussed; there exists very little discriminatory power in this data which would render more sophisticated analysis such as structural equation modelling unlikely to be able to reveal further insights.

3. Modelling methodology

The data format for BWS case 1 is similar to that of an unlabelled choice experiment. The alternatives, denoted *j*, in effect reflect a specific position within the task (e.g., 1 = top, 2 = second from the top, 3 = second from the bottom, 4 = bottom), whilst the objects, in this case the various statements, are represented as the attributes. For each task, *s*, two observations are constructed, one reflecting the best choice, and a second pseudo observation representing the worst choice. For the best choice observations, each statement is dummy coded 1 if it is present in alternative, *j*, or 0 otherwise. For the worst choice task, the variables are simply the negative of the best values (i.e., −1, if the alternative is present or 0 otherwise). In the current context, the fact that there exist 25 statements results in the generation of 24 dummy coded variables. As is common practice with such data, the alternative chosen as best was removed from the set of set available attributes (statements) for estimation of the second response; the attribute chosen as worst (i.e., it is assumed that respondents choose the best from all available attributes, then choose the worst from the now limited remaining set of candidate attributes).

Given the presence of both satisfaction and importance choices, two utility specifications are present in the current model set up. Let $V_{(\mathrm{Sat})nsj}$ and $V_{(\mathrm{Imp})nsj}$ denote the utility for the satisfaction and importance questions associated with alternative *j* as perceived by respondent *n* in choice task *s*, respectively. For both utility functions, a linear in the parameters specification is assumed. In the current paper, we further assume that the marginal utilities for the best and worst choices are symmetrical, however we allow for the possibility that there exist scale differences between the two choices (see e.g., Dyachenko et al., 2014; Scarpa et al., 2011; Rose, 2014) by estimating scale parameters, $\lambda_{(\cdot)worst}$, associated with the worst choice tasks (i.e., $\lambda_{(\cdot)worst}$ is normalized to zero for the best choice tasks). Let δ_{nk} denote the marginal utility for satisfaction associated with

Table 4
p-values for differences in mean for satisfaction and importance ratings scales.

Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1		1.000	1.000	1.000	1.000	0.315	0.315	1.000	1.000	0.112	1.000	1.000	0.850	1.000	1.000	0.998	0.980	1.000	0.412	1.000	1.000	1.000	1.000	1.000	1.000
2	1.000		0.992	0.989	0.630	**0.004**	**0.004**	1.000	0.930	**0.001**	1.000	0.630	0.064	0.766	0.989	0.412	0.207	0.557	**0.007**	0.734	0.895	1.000	1.000	1.000	0.956
3	0.985	0.284		1.000	1.000	0.630	0.630	0.998	1.000	0.315	1.000	1.000	0.980	1.000	1.000	1.000	0.999	1.000	0.734	1.000	1.000	1.000	1.000	1.000	1.000
4	1.000	0.795	1.000		1.000	0.666	0.666	0.998	1.000	0.346	1.000	1.000	0.985	1.000	1.000	1.000	1.000	1.000	0.766	1.000	1.000	1.000	1.000	1.000	1.000
5	1.000	0.985	1.000	1.000		0.992	0.992	0.766	1.000	0.913	0.989	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998	1.000	1.000	0.965	1.000	0.985	1.000
6	1.000	1.000	0.518	0.943	0.999		1.000	**0.009**	0.874	1.000	0.086	0.992	1.000	0.974	0.666	0.999	1.000	0.996	1.000	0.980	0.913	**0.048**	0.346	0.074	0.824
7	1.000	1.000	0.183	0.665	0.955	1.000		**0.009**	0.874	1.000	0.086	0.992	1.000	0.974	0.666	0.999	1.000	0.996	1.000	0.980	0.913	**0.048**	0.346	0.074	0.824
8	0.995	0.377	1.000	1.000	1.000	0.629	0.256		0.974	**0.002**	1.000	0.766	0.112	0.874	0.998	0.557	0.315	0.701	**0.015**	0.850	0.956	1.000	1.000	1.000	0.985
9	1.000	0.973	1.000	1.000	1.000	0.998	0.929	1.000		0.594	1.000	1.000	0.999	1.000	1.000	1.000	1.000	1.000	0.930	1.000	1.000	0.999	1.000	1.000	1.000
10	1.000	1.000	0.996	1.000	1.000	1.000	0.998	0.999	1.000		**0.022**	0.913	1.000	0.824	0.346	0.980	0.998	0.944	1.000	0.850	0.666	**0.011**	0.128	**0.018**	0.520
11	1.000	0.980	1.000	1.000	1.000	0.998	0.943	1.000	1.000	1.000		0.989	0.483	0.998	1.000	0.944	0.796	0.980	0.128	0.996	1.000	1.000	1.000	1.000	1.000
12	1.000	0.913	1.000	1.000	1.000	0.985	0.823	1.000	1.000	1.000	1.000		1.000	1.000	1.000	1.000	1.000	1.000	0.998	1.000	1.000	0.965	1.000	0.985	1.000
13	1.000	0.765	1.000	1.000	1.000	0.929	0.629	1.000	1.000	1.000	1.000	1.000		1.000	0.985	1.000	1.000	1.000	1.000	1.000	1.000	0.346	0.874	0.447	0.998
14	0.985	0.284	1.000	1.000	1.000	0.518	0.183	1.000	1.000	0.996	1.000	1.000	1.000		1.000	1.000	1.000	1.000	0.989	1.000	1.000	0.989	1.000	0.996	1.000
15	0.973	0.230	1.000	1.000	1.000	0.446	0.144	1.000	1.000	0.992	1.000	1.000	1.000	1.000		1.000	1.000	1.000	0.766	1.000	1.000	1.000	1.000	1.000	1.000
16	0.344	**0.007**	1.000	0.985	0.795	**0.025**	**0.003**	1.000	0.849	0.482	0.823	0.943	0.989	1.000	1.000		1.000	1.000	1.000	1.000	1.000	0.874	0.999	0.930	1.000
17	0.482	**0.015**	1.000	0.996	0.894	**0.047**	**0.007**	1.000	0.929	0.629	0.913	0.980	0.998	1.000	1.000	1.000		1.000	1.000	1.000	1.000	0.666	0.985	0.766	1.000
18	0.206	**0.003**	1.000	0.943	0.629	**0.011**	**0.001**	0.999	0.699	0.313	0.665	0.849	0.955	1.000	1.000	1.000	1.000		0.999	1.000	1.000	0.944	1.000	0.974	1.000
19	0.733	**0.047**	1.000	1.000	0.980	0.127	**0.025**	1.000	0.989	0.849	0.985	0.998	1.000	1.000	1.000	1.000	1.000	1.000		0.992	0.956	0.074	0.447	0.112	0.895
20	0.999	0.555	1.000	1.000	1.000	0.795	0.411	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	1.000	0.992	1.000		1.000	0.985	1.000	0.995	1.000
21	0.849	0.085	1.000	1.000	0.995	0.206	**0.047**	1.000	0.998	0.929	0.996	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000		0.998	1.000	1.000	1.000
22	1.000	0.999	0.998	1.000	1.000	1.000	0.996	0.999	1.000	1.000	1.000	1.000	1.000	0.998	0.995	0.518	1.000	0.344	0.873	1.000	0.943		1.000	1.000	1.000
23	0.765	0.055	1.000	1.000	0.985	0.144	**0.030**	1.000	0.992	0.873	0.989	0.999	1.000	1.000	1.000	1.000	0.665	1.000	1.000	1.000	1.000	0.894		1.000	1.000
24	1.000	0.973	1.000	1.000	1.000	0.998	0.929	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.849	1.000	0.699	0.989	1.000	0.998	1.000	0.992		1.000
5	0.699	**0.041**	1.000	1.000	0.973	0.112	**0.021**	1.000	0.985	0.823	0.980	0.998	1.000	1.000	1.000	1.000	0.929	1.000	1.000	1.000	1.000	0.849	0.985	0.985	

Bolded font represent insignificant results at the 5% level.

the kth statement and similarly θ_{nk} the marginal utility for importance associated with same statement. Alternative specific constants are estimated for the first three alternatives in both cases, which reflect positional or other order effects that might exist within the data. In addition to the observed component of utility, both models are assumed to have an error term. The utility functions thus described are provided in Eqs. (1) and (2).

$$V_{(\text{Sat})nsj} = \exp(\lambda_{(\text{Sat})worst})\left(\alpha_{(\text{Sat})j} + \sum_{k=1}^{K=24}\delta_{nk}x_{nsjk}\right) + \xi_{(\text{Sat})nsj}, \tag{1}$$

$$V_{(\text{Imp})nsj} = \exp(\lambda_{(\text{Imp})worst})\left(\alpha_{(\text{Imp})j} + \sum_{k=1}^{K=24}\theta_{nk}x_{nsjk}\right) + \xi_{(\text{Imp})nsj}. \tag{2}$$

The error terms of the two specifications are assumed to be Normally distributed with mean zero, and covariance Ω_e. Given that respondents answer both the satisfaction and importance tasks simultaneously, it is plausible that the error terms of the two are correlated. In order to account for such a possibility, the model allows for the estimation of a correlation term, ρ, which is designed to uncover the degree of correlation, if any, between the error terms of the satisfaction and importance utility functions. For purposes of model identification, we normalise the variances of the random error terms in both models to 1.0. The joint error structure of the model is given as Eq. (3).

$$\begin{bmatrix}\xi_{(\text{Sat})nsj}\\ \xi_{(\text{Imp})nsj}\end{bmatrix} \sim N\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}1 & \rho\\ \rho & 1\end{bmatrix}\right), \ -1 \leqslant \rho \leqslant 1. \tag{3}$$

To estimate the error structure as described in Eq. (3), we rely on Cholesky decomposition, which is shown in Eq. (4). The specific Choleksy matrix used in this context is the lower triangular matrix described by the first matrix in the right hand side of Eq. (4).

$$\begin{bmatrix}1 & \rho\\ \rho & 1\end{bmatrix} = \begin{bmatrix}1 & 0\\ \rho & \sqrt{1-\rho^2}\end{bmatrix}\begin{bmatrix}1 & \rho\\ 0 & \sqrt{1-\rho^2}\end{bmatrix} \tag{4}$$

where draws from the bivariate distribution are obtained from independent $N(0,1)$ draws, r_{1nsj} and r_{2nsj} such that

$$\begin{aligned} r_{(\text{Sat})nsj} &= r_{1nsj},\\ r_{(\text{Imp})nsj} &= r_{1nsj}\rho + r_{2nsj}\sqrt{1-\rho^2}. \end{aligned} \tag{5}$$

The correlation parameter ρ is constant across all j alternatives, however it is identified due to the fact that differences are generated via the independent normal draws, r_{1nsj} and r_{2nsj} (see Hess et al., 2008). Given that the error terms are assumed to be normally distributed, the resulting model is a bivariate Multinomial Probit model (see Abou Zeid, 2009 as discussed in Bierlaire and Fetiarison, 2009).

The satisfaction and importance parameter estimates in Eqs. (1) and (2) are assumed to be randomly distributed over the population following normal distributions. Rather than assume univariate normal distributions, the model allows for correlations between the importance and satisfaction parameter estimates related to the same statement, k, but does not allow for correlation of the random parameters between statements. Whilst it is behaviourally plausible that the marginal utilities for the various statements are correlated, a model allowing for just such a multivariate distribution would require the estimation of 600 parameters (i.e., the Cholesky matrix for the satisfaction and importance parameters each require the estimation of 300 terms each), making the model econometrically intractable. Eq. (6) represents the structural random parameter terms estimated for the kth statement, where μ_{k1} and μ_{k2} represent the means of the satisfaction and importance distributions respectively, and σ_{k1} and σ_{k2} the variance terms. σ_{k1k2} in Eq. (6) represents the covariance between the two distributions.

$$\begin{bmatrix}\delta_{nk}\\ \theta_{nk}\end{bmatrix} \sim N\left(\begin{bmatrix}\mu_{k1}\\ \mu_{k2}\end{bmatrix}, \begin{bmatrix}\sigma_{k1} & \sigma_{k1k2}\\ \sigma_{k1k2} & \sigma_{k2}\end{bmatrix}\right). \tag{6}$$

Rather than estimate σ_{k1}, σ_{k2} and σ_{k1k2} directly, the model involves the estimation of S_{11k}, S_{22k} and S_{12k} derived from the Cholesky decomposition matrix of the covariance structure of Eq. (6), as shown in Eq. (7).

$$\Omega_k = \left(\begin{bmatrix}S_{11k} & 0\\ S_{12k} & S_{22k}\end{bmatrix}\begin{bmatrix}S_{11k} & S_{12k}\\ 0 & S_{22k}\end{bmatrix}\right). \tag{7}$$

Given the model structure as described above, we use a logit-smoothed AR simulator (see Train, 2009) to approximate the probability that respondent n is observed to choose alternative j in choice task s as being the statement they are most or least satisfied such that

$$P_{(\text{Sat})nsj} = \int_{\delta}\int_{\xi_{(\text{Sat})}} \frac{\exp(V_{(\text{Sat})nsj})}{\sum_{i\in J_{ns}}\exp(V_{(\text{Sat})nsi})} f(\delta|\Omega_p(\mu,\sigma)) f(\xi_{(\text{Sat})}|\Omega_e) d\delta d\xi_{(\text{Sat})}, \tag{8}$$

where $f(\delta|\Omega_p(\mu,\sigma))$ is the multivariate probability density function of δ, given the distributional parameters $\Omega_p(\mu,\sigma)$ and $f(\xi_{(\text{Sat})}|\Omega_e)$ is the multivariate probability density function of the joint error terms between the satisfaction and importance questions, described by Eq. (3).

Similarly, the probability that respondent n is observed to choose alternative j in choice task s as being the statement they are most or least important with is

$$P_{(\text{Imp})nsj} = \int_{\theta}\int_{\xi_{(\text{Imp})}} \frac{\exp(V_{(\text{Imp})nsj})}{\sum_{i\in J_{ns}}\exp(V_{(\text{Imp})nsi})} f(\theta|\Omega_p(\mu,\sigma))f(\xi_{(\text{Imp})}|\Omega_e)d\theta d\xi_{(\text{Imp})}, \tag{9}$$

where $f(\theta|\Omega_p(\mu,\sigma))$ is the multivariate probability density function of θ, given the distributional parameters $\Omega_p(\mu,\sigma)$ and $f(\xi_{(\text{Imp})}|\Omega_e)$ is the multivariate probability density function of the joint error terms between the satisfaction and importance questions. Note that Ω_p appears in both Eqs. (8) and (9) to reflect the fact that the two are correlated, as described by Eq. (6).

The joint probability that respondent n will be observed to make a sequence of choices S_n choices $\{j_{(\text{Sat})}|y_{(\text{Sat})nsj}=1\}_{s\in S_n}$ and $\{j_{(\text{Imp})}|y_{(\text{Imp})nsj}=1\}_{s\in S_n}$ for both the satisfaction and importance questions is given as by

$$P_n = \int_{\delta}\int_{\theta}\int_{\xi_{\text{Sat}}}\int_{\xi_{\text{Imp}}} \prod_{s\in S_n}\prod_{j\in J_{ns}} (P_{(\text{Sat})nsj})^{y_{(\text{Sat})nsj}} (P_{(\text{Imp})nsj})^{y_{(\text{Imp})nsj}} f(\delta|\Omega_p(\mu,\sigma))f(\theta|\Omega_p(\mu,\sigma))f(\xi_{(\text{Sat})}|\Omega_e)f(\xi_{(\text{Imp})}|\Omega_e)d\delta d\theta d\xi_{(\text{Sat})}d\xi_{(\text{Imp})}. \tag{10}$$

The log-likelihood function for the model based on Eq. (10) is

$$LL = \sum_n \ln(P_n). \tag{11}$$

Given that the probabilities obtained from Eqs. (8)–(10) are not of a closed form, the model is estimated using simulated maximum likelihood (see e.g., Train, 2009).

4. Empirical results

Table 5 presents the results of the modelling exercise based on the best–worst task. The model was estimated using Python-Biogeme version 2.4 (Bierlaire, 2003) using the CFSQP algorithm and 2500 modified Latin hypercube sampling (MLHS) draws (see Hess et al., 2005). For both the satisfaction and importance results, the last item (i.e., shelter provided at bus stop at end of journey) was treated as the base of the dummy coding scheme. As such, it is important to note that parameters that are statistically insignificant are neither unimportant or sources of dissatisfaction, but rather are not statistically different to the base statement. This is particularly important given that several parameters were found to be statistically significant and negative relative to the base, and hence also relative to any parameter that is statistically insignificant in the model. The table is reported in several sections. First reported are the satisfaction results. Shown are the mean and Cholesky parameters associated with the normal distributions estimated for each of the dummy coded items. Note that the Cholesky term for satisfaction represents the standard deviation parameter for the item.

Next, the results for the importance choices are shown. Again, reported at the means and Cholesky parameters related to the Normal distributions obtained from the modelling process. Unlike the satisfaction results, two Cholesky terms are reported for the importance results, the diagonal and off-diagonal terms of the Cholesky upper triangular matrix. Note that the standard deviation for the Normal distribution associated with each importance item, can be computed as

$$\sqrt{(C_{diag.})^2 + (C_{off\text{-}diag.})^2}. \tag{12}$$

The standard deviation parameters thus calculated are presented next to the mean parameter estimates in italics. Standard errors for the standard deviation parameters are computed using the Delta method (see Greene, 2007). The standard deviations for the satisfaction questions may be interpreted as the diagonal of the Cholesky associated with the satisfaction questions. The final column of the table reports the pairwise correlation for the same item between the satisfaction and importance choices. The correlation is computed by converting the item specific Cholesky matrix to a covariance matrix, from which the correlation can be derived.

For both the satisfaction and importance utility functions, three ASCs were estimated to detect and correct for any positional biases that might exist within the data. In both instances, the first and second ASCs were found to be statistically significant and positive suggesting that respondents were more likely to select the top two items shown in the task, *all else being equal*. For the importance questions, the third ASC was also positive and statistically significant suggesting that this item was also chosen more frequently than the last item shown in the task, *all else being equal*. For the satisfaction questions, the ASC for the third alternative was found not to be statistically significant suggesting that this item was selected no more or less than the last item shown in each task, *ceteris paribus*.

With regards to the mean satisfaction results, on average, 12 items were found to be statistically no different to the base and hence respondents share a similar level of satisfaction with these attributes, whilst eight of the parameters are statistically significant and positive, indicating that respondents are significantly more satisfied with these attributes relative to

Table 5
Model results (base attribute = shelter provided at bus stop at the end of journey).

	Satisfaction				Importance						Correlation term		
	Par.	(rob. *t*-rat.)	Par.	(rob. *t*-rat.)	Par.	(rob. *t*-rat.)	Par.	(rob. *t*-rat.)	Par.	(rob. *t*-rat.)	Par.	(rob. *t*-rat.)	
	Mean		Cholesky diagonal		Mean		Est. std dev.		Cholesky diagonal		Cholesky off-diagonal		Correl.
Alternative specific constants (order effects)													
Alternative specific constant 1	0.270	(−2.82)	–	–	0.753	(8.98)	–	–	–	–	–	–	–
Alternative specific constant 2	0.210	(2.44)	–	–	0.338	(3.97)	–	–	–	–	–	–	–
Alternative specific constant 3	−0.075	(−0.85)	–	–	0.273	(3.23)	–	–	–	–	–	–	–
Statement effects													
Information available on website	0.416	(2.09)	−1.030	(−3.70)	0.953	(4.99)	−0.959	(−4.55)	0.997	(5.02)	−0.274	(−1.87)	0.274
Location of bus stop (at start of journey)	0.607	(3.31)	−0.544	(−2.52)	0.927	(5.07)	−0.340	(−1.44)	0.384	(0.89)	−0.178	(−1.37)	0.463
Cleanliness at bus stop (at start of journey)	0.015	(0.08)	0.409	(2.28)	−0.472	(−2.49)	0.103	(0.21)	0.435	(0.63)	0.423	(1.64)	0.972
Signage at bus stop (at start of journey)	0.150	(0.79)	−0.875	(−4.08)	0.299	(1.56)	−0.514	(−2.50)	0.907	(2.37)	−0.747	(−3.86)	0.824
Shelter provided at bus stop (at start of journey)	0.021	(0.11)	−0.321	(−1.49)	−0.024	(−0.12)	−0.672	(−2.95)	1.168	(3.01)	0.955	(4.31)	−0.818
Frequency of bus service	−0.268	(−1.33)	−1.000	(−4.42)	1.570	(7.86)	0.965	(3.45)	1.062	(3.28)	0.643	(2.24)	−0.555
Bus being on time (keeping to timetable)	−0.544	(−2.69)	0.757	(2.60)	2.000	(9.93)	0.591	(2.49)	1.063	(8.61)	−0.884	(−4.70)	−0.831
Ease of boarding the bus	0.467	(2.55)	0.585	(2.50)	−0.350	(−1.75)	−1.080	(−5.24)	1.099	(2.42)	0.202	(0.82)	0.184
Ease of payment	0.434	(2.36)	−0.458	(−1.29)	0.286	(2.50)	−0.893	(−3.71)	0.898	(2.89)	0.972	(2.49)	−0.736
Cost of bus trip	−0.201	(−2.91)	1.560	(5.83)	1.000	(5.16)	−1.180	(−4.81)	1.192	(2.48)	−0.169	(−0.78)	−0.142
Cleanliness of the bus	0.375	(2.06)	0.466	(2.14)	0.346	(1.97)	−0.282	(−1.60)	0.293	(0.74)	−0.078	(−0.38)	−0.267
Friendliness of staff	0.442	(2.48)	0.115	(0.44)	0.627	(3.36)	0.205	(0.95)	0.666	(3.35)	−0.634	(−2.70)	−0.951
Knowledgeability of staff	−0.578	(−2.97)	0.550	(2.56)	0.178	(0.98)	0.227	(0.91)	0.296	(0.67)	−0.190	(−0.88)	−0.642
Level of crowding	−0.186	(−0.97)	−0.691	(−3.16)	0.405	(2.18)	0.056	(0.23)	0.783	(2.92)	0.781	(3.87)	−0.997
Ability to get a seat	0.121	(0.67)	−0.171	(−0.88)	−0.146	(−0.72)	0.436	(2.05)	0.940	(3.60)	0.833	(3.77)	−0.886
Comfort of seat	−0.098	(−0.52)	−0.624	(−2.22)	−0.425	(−2.10)	0.723	(2.62)	1.069	(2.09)	−0.788	(−2.82)	0.737
Leg room on bus	−0.323	(−1.55)	−1.010	(−4.03)	−0.085	(−0.46)	−0.215	(−0.97)	0.510	(1.09)	0.463	(1.98)	−0.907
Comfort of ride	−0.233	(−1.25)	0.106	(0.56)	−1.010	(−4.55)	0.780	(3.26)	1.655	(4.52)	1.460	(4.59)	0.882
Noise on bus	−0.504	(−2.53)	0.674	(2.56)	−0.601	(−3.16)	−0.324	(−1.27)	0.342	(0.78)	0.111	(0.49)	0.324
Travel time	−0.225	(−1.14)	−0.870	(−4.16)	0.114	(0.60)	0.650	(3.05)	0.655	(4.09)	0.977	(3.43)	−0.832
Temperature	0.025	(0.14)	−0.117	(−0.64)	0.891	(4.43)	−0.247	(−1.15)	1.225	(24.39)	1.200	(4.37)	−0.979
Location of bus stop (at end of journey)	0.732	(4.07)	0.186	(0.97)	0.913	(4.87)	0.103	(0.37)	0.455	(2.28)	0.443	(2.26)	0.974
Cleanliness at bus stop (at end of journey)	−0.105	(−0.54)	0.707	(3.08)	−0.325	(−1.68)	0.484	(2.65)	0.725	(1.83)	0.540	(2.11)	0.745
Ease of connection to next service (at end of journey)	0.186	(2.00)	0.371	(1.91)	0.097	(0.47)	0.143	(0.66)	1.238	(3.18)	1.230	(5.84)	0.993
Shelter provided at bus stop (at end of journey)	0.00	–	–	–	0.00	–	–	–	–	–	–	–	–
Scale term													
exp(Scale Worst)	−0.574	(−7.63)	–	–	−0.333	(−5.19)	–	–	–	–	–	–	–
Error term correlation													
Error term correlation	–	–	–	–	–	–	–	–	–	–	−0.635	(−18.79)	−0.635
Model fit													
LL(0)	−9989.325												
LL(6)	−9089.698												
ρ^2	0.090												
Adj. ρ^2	0.060												

Item statement	Satisfaction	Importance	Correlation
Beginning of journey			
Information available on website	75.88	65.22	0.27
Location of bus stop (at start of journey)	90.46	64.35	0.46
Cleanliness at bus stop (at start of journey)	45.27	17.87	0.97
Signage at bus stop (at start of journey)	55.57	43.49	0.82
Shelter provided at bus stop (at start of journey)	45.69	32.74	-0.82
On the bus			
Frequency of bus service	23.66	85.71	-0.55
Bus being on time (keeping to timetable)	2.60	100.00	-0.83
Ease of boarding the bus	79.77	21.93	0.18
Ease of payment	77.25	43.06	-0.74
Cost of bus trip	28.78	66.78	-0.14
Cleanliness of the bus	72.75	45.05	-0.27
Friendliness of staff	77.86	54.39	-0.95
Knowledgeability of staff	0.00	39.47	-0.64
Level of crowding	29.92	47.01	1.00
Ability to get a seat	53.36	28.70	-0.89
Comfort of seat	36.68	19.44	0.74
Leg room on bus	19.47	30.72	-0.91
Comfort of ride	26.34	0.00	0.88
Noise on bus	5.65	13.59	0.32
Travel time	26.95	37.34	-0.83
Temperature	46.05	63.16	-0.98
At the end of the journey			
Location of bus stop (at end of journey)	100.00	63.89	-0.98
Cleanliness at bus stop (at end of journey)	36.11	22.76	0.97
Ease of connection to next service (at end of journey)	58.32	36.77	0.74
Shelter provided at bus stop (at end of journey)	44.12	33.55	0.99

Fig. 2. Best worst relative importance and satisfaction results.

the others.[4] The remaining four parameters are statistically significant and negative suggesting that satisfaction is significantly lower over these attributes relative to the others. Given that all items are dummy coded and hence are coded using the same unit of measure, it is possible to directly compare the parameter estimates in terms of their magnitude. On average, the item resulting in the greatest level of satisfaction appears to be the location of the bus stop at the end of the journey followed by the location of the bus stop at the commencement of the journey, ease of payment, and the friendliness of the staff. The items that generate the least satisfaction for the survey sample are how knowledgeable the staff are, the ability of a bus to keep to time, and the noise of the bus.

For the mean importance component of the scale, on average, nine items were found to be statistically no different to the base level, with 11 items considered statistically as being more important than whether shelter is provided at a bus stop at end of journey, and four items as being statistically of less importance. Of greatest importance to the sampled respondents is the bus being on time followed by the frequency of the bus service, and the cost of the trip. Information about the trip being provided on the internet, the location of the bus stop at the commencement of the journey, the on-board temperature, and the location of the bus stop at the end of the journey are also of huge import to the respondents. With regards to the bus being on-time, it is clear that this is important to the respondents as an attribute, however the same respondents are not satisfied with this aspect of their journey. Of least importance to the respondents are the comfort of the bus ride and the noise of the bus during the journey.

The last column of the table presents the correlation coefficient between the satisfaction and importance parameters. Negative coefficients suggest that respondents who derive positive satisfaction for an item tend to not believe that same item to be important. Inversely, respondents who are less satisfied with the same item, tend to believe that that item is

[4] Note that changing the base statement to the statement that has the largest negative parameter estimate would simply make all the items shown statistically significant and positive relative to the new base, however the items that are statistically no different to the current base would not be statistically significantly different to one another.

Item statement	Satisfaction	Importance	Correlation
Beginning of journey			
Information available on website	72.12	69.31	0.54
Location of bus stop (at start of journey)	100.00	96.04	0.47
Cleanliness at bus stop (at start of journey)	63.46	29.70	0.36
Signage at bus stop (at start of journey)	62.50	44.55	0.38
Shelter provided at bus stop (at start of journey)	45.19	56.44	0.32
On the bus			
Frequency of bus service	8.65	89.11	0.31
Bus being on time (keeping to timetable)	8.65	100.00	0.23
Ease of boarding the bus	96.15	32.67	0.45
Ease of payment	55.77	54.46	0.31
Cost of bus trip	0.00	65.35	0.10
Cleanliness of the bus	82.69	55.45	0.35
Friendliness of staff	45.19	49.50	0.35
Knowledgeability of staff	24.04	43.56	0.24
Level of crowding	49.04	29.70	0.35
Ability to get a seat	62.50	27.72	0.20
Comfort of seat	39.42	4.95	0.33
Leg room on bus	32.69	8.91	0.35
Comfort of ride	43.27	0.00	0.30
Noise on bus	11.54	15.84	0.36
Travel time	48.08	37.62	0.32
Temperature	53.85	19.80	0.46
At the end of the journey			
Location of bus stop (at end of journey)	86.54	64.36	0.46
Cleanliness at bus stop (at end of journey)	71.15	16.83	0.33
Ease of connection to next service (at end of journey)	83.65	54.46	0.39
Shelter provided at bus stop (at end of journey)	57.69	14.85	0.32

Fig. 3. Traditional scale relative importance and satisfaction results.

of higher importance. For positive correlation coefficients, the reverse interpretation can be made. Thus for example, despite being statistically significant and positive for both importance and satisfaction, the negative correlation of −0.951 for the item '*friendliness of staff*' suggests that those respondents who are satisfied with this aspect of their service are not likely to believe that this is an important service component, whilst those who are least satisfied with this service component, are more likely to rate it as being important to them. In terms of '*Location of bus stop (at end of journey)*', the positive correlation coefficient of 0.974 suggests that those respondents who believe this is an important component of the journey are more likely to be satisfied with this attribute, whilst those who care very little about how friendly the staff are, are more likely to be less satisfied with this aspect of their travel experience. Of note, ten of the 24 correlations reported are less than −0.5, with eight being greater than 0.5. This suggests somewhat of a disconnect between the degree of satisfaction and importance experienced by most respondents with most respondents experiencing less satisfaction for a greater number of service components they find important, then they experience satisfaction for service aspects they find important. What should be of significant concern to the transit operators of the services used by the sample population is the fact that for the most important component 'Bus being on time (keeping to timetable)', is that this service component has on average a negative satisfaction result, and there exists a negative and quite large correlation between the two questions.

Examining the scale parameters associated with the worst questions, the parameters for both satisfaction and importance are statistically significant and negative suggesting that there exists greater error variance within the sample when providing the worst responses relative to the best responses. The results suggest that the magnitude of the worst parameters for the satisfaction questions are 0.563 times those of the best, and 0.716 times that of the best questions for the importance questions. This is a similar finding to others using the best–worst paradigm, with the (first) worst choices tending to produce

greater error variances than the (first) best (e.g., Dyachenko et al., 2014; Scarpa et al., 2011; Rose, 2014). Finally, also reported is the correlation coefficient for the joint satisfaction and importance error term. The error term correlation is statistically significant and negative, suggesting that all else being equal, there is a negative relationship between any omitted items in terms of the degree of satisfaction and importance experienced.

To further aid in interpreting the results, Fig. 2 plots rescaled estimates from the model reported in Table 5. Rescaling is achieved by taking the difference between the parameter estimate for each item and the minimum estimate from the model, and dividing the result by the parameter range (i.e., the difference between the largest and smallest magnitude parameters). This rescaling procedure converts the estimates to a zero-one scale where the least important (or satisfying) item will be given a zero and the most important (or satisfying) item will be given a value of one. Note that this is based solely on the means of the Normal distributions shown in Table 5, and hence does not account for any heterogeneity observed within the sample. In green are the rescaled results for the importance index, whilst the satisfaction index is shown in blue. Differences between the two columns for the same item indicate a discrepancy in terms of the relative satisfaction and importance for that item. For comparison we also provide the rescaled data for the traditional ratings scale data using the same approach in Fig. 3.

Comparing the two figures, in the BWS data is can be clearly seen that there are two service attributes that particularly matter (being on time and of sufficient frequency), whereas in Fig. 3 these two plus the location of the bus stop at the start of the journey are at the top. In terms of what respondents are least satisfied with, the BWS results in Fig. 2 clearly reveal knowledgeability of staff, bus being on time and noise on the bus as being the least satisfied attributes. On the other hand, the traditional results presented in Fig. 3 reveal cost, frequency, running to timetable and noise on the bus to be the least satisfying aspect of bus use. It could be argued that dissatisfaction with these items would be what one would expect to uncover when surveying public transport users. The BWS method arguably delivers a better insight in that it forces respondents to think about relative levels of satisfaction, thus requires greater differentiation than responding that cost and reliability are "bad". Perhaps the most striking result that can be seen is that all correlations are positive in Fig. 3 which, as discussed previously, may be an artefact of the way in which data is collected using a multi-item scale such as the one deployed in this study.

5. Discussion and conclusion

In this paper, we demonstrate how a dual response best worst experiment can be used to generate a service quality index for public transport usage. The dual satisfaction and importance response task proposed and implemented here allowed for the first time in the wider literature (to the knowledge of the authors), the recovery of the correlation between satisfaction and importance in attitudes. Significantly, the modelling revealed several such correlations. The importance of such a finding is that it allows transit authorities the ability not just to understand the average impact each item of the index has on satisfaction and importance, but how the two move together. Armed with such knowledge, transit authorities will be better capable of understanding which items should be of greater concern to them in terms of any attempts to improve the quality of a passenger's journey.

The paper also presented the results of a traditional ratings task covering the same items used in the best worst task. The results of the ratings task demonstrated clearly difficulties in using such survey questions for policy and/or decision makers. As shown, very little discernible differences were found to exist in terms of how respondents rated their level of satisfaction and importance with each of the 25 items, meaning that for a transit operator, it is not clear which item would be best targeted to improve the quality of their service. In contrast, the best worst task was clearly able to distinguish both in terms of satisfaction and importance, differences between each of the items. In this manner, the method provides clearly actionable information for transit operators, as well as regulators. It is worth noting that this method can easily be applied to wider attitudinal studies.

5.1. Comparing the approaches

One important way in which the best–worst approach proposed in this paper differentiates itself from traditional ratings scales is in the examination of the correlations between importance and satisfaction. With respect to the ratings scales, all the correlations between importance and satisfaction are positive; suggesting that for each item, greater (lower) reported importance is concomitant with higher (lower) levels of reported satisfaction. Whilst it is not possible to rule out that such relationships may truly exist in practice, it is highly improbable that respondents (bus users) are satisfied with all important aspects of their journey. We hypothesise that respondents undertaking this survey did not fully engage with the rating scale items either due to their repetitiveness or due to the number of items shown. Consequently they simple rate every item as having equal importance or equal satisfaction. For example, in a list of 25 service factors, every item may be given either a four or five for importance and a three or four for satisfaction. The end result is a scale with little differentiation and largely positive correlations as observed herein. On the other hand, the best–worst approach presented in this paper is able to recover both positive and negative correlation structures between the importance and satisfaction questions. Presenting the survey in the best–worst format may therefore be more discriminating in that requires that respondents consider more closely what they feel is the most important service offering and/or what they are the most or least satisfied with. With

respondents more engaged, the survey approach better replicates the differentiation in importance and satisfaction that is likely to exist, by not allowing the respondent to simply say everything is the same. We acknowledge that this finding may be specific to this particular survey.

Comparing the results obtained from each approach, the five most important aspects of bus travel obtained from the traditional ratings task are the bus being on time, the location of the stop at the start of the trip, information available on the website, the location of the stop at the end of the trip and the cost of the journey. While these are the attributes with the highest average ratings, statistically the averages for each individual item are the same not only among these five, but also among many of the remaining service attributes. Consequently, a transport operator would not be confident that these five items really are the most important from amongst the set and hence would be uncertain in practice as to which aspects of bus trips under their control they should target for potential changes. Additionally, all of these service attributes are positively correlated with satisfaction suggesting that the more important an item is rated, the more likely bus users are to be satisfied with that attribute; a convenient result for transit operators in that customers are satisfied with everything they find important. Subject to similar results being found to exist across multiple bus operators, findings of this nature would be expected to have significant implications for transit regulators attempting to undertake benchmarking type exercises. Any regulator attempting to compare customer satisfaction will have difficulty in discerning the relative importance and satisfaction of various operators in the market, as well as determining what if any service characteristics should be improved across the network.

In contrast, based on the best–worst approach introduced here, the five most important service attributes are the bus being on time, the frequency of service, the cost of the trip, information on the website and location of the stop at the start of the journey. Not only does this method produce a different ordering of attributes, immediately it can be seen that the bus being on time and the frequency of service are the two most important service characteristics by some margin. Additionally, the negative correlation between importance and satisfaction for both these attributes is informative to both transit operators and regulators suggesting that whilst these trip characteristics are important to transit users, they are not satisfied with these attributes. Unlike the ratings scale approach where all service characteristics are deemed equally important and engender equal amounts of satisfaction, the information gained from the best–worst approach provides clear operational guidance to operators and regulators in terms of what aspects of transit trips should be examined more closely in order to improve customer service in the future. The fact that this is an important operational issue is more clearly highlighted in the best–worst approach than in a ratings scale approach whereby everything is important and everything engenders equal satisfaction.

One acknowledged disadvantage of the best–worst approach is that it requires the use of a more complex survey design which would require training of operators and regulators to implement. We note further that we have used a relatively complex model form in this paper, however simpler models such as the multinomial logit model could be used, at the loss of some information (such as the correlation structures). From a practical perspective, a further possible approach would be to obtain a significantly large representative sample from which the more complex model could be estimated, after which, respondents surveyed in subsequent time periods could be given the best–worst survey task, and their answers feed through the pre-estimated model. In doing so, it would be necessary to assume that the weights assigned to importance and satisfaction are stable over time, however this could be tested by periodically updating the estimated model.

5.2. Policy implications of the analysis

With respect to how a public transport authority or operator might action the results from our proposed methodology, the most immediate place to start would be to identify the attributes that respondents identify as being most important and assess the relative level of satisfaction they have with those service attributes. In our study we have revealed running to timetable and frequency of service as being the most important attributes to bus users and that clearly there is low relative satisfaction with buses running on time marking it as a particularly critical attribute for improving service levels. In the case of frequency of service, the relative level of satisfaction is low, but there are other service attributes that are viewed as being worse. Looking at the factors that respondents are least satisfied with, it is apparent that these attributes are noise on the bus and knowledge of staff, however relatively speaking noise on the bus is one of the least important of the 25 service attributes marking it as a point of concern, but perhaps not an attribute that needs immediate attention. Interestingly in the best worst approach, cost is a relatively middle of the road attribute in terms of both importance and level of satisfaction, suggesting that for operators and transit authorities that there are other factors of public transport trips which should be targeted first for improvement. Some final insights from the results that might of interest to those in public transportation are revealed by examining the correlations. The level of crowding, cleanliness of the bus stops at both ends of the journey, as well as having shelter at the end of the trip all exhibit large positive correlations between satisfaction and importance indicating that as bus users find these variables to be more important they are more satisfied with them (or vice versa). This suggests that crowding may be something that needs to be emphasised by operators and that perhaps further research may be needed into what constitutes crowding and why those who find it important are satisfied with the level of crowding.

On the other hand, some of the negative correlations are quite striking and perhaps offer novel insights. For example, for respondents who are find the temperature on the bus to be relatively more important, there are lower relative levels of satisfaction. This marks on-board temperature as something that bus service providers may wish to monitor, particularly in more extreme climates or on particularly hot or cold days, but more generally noting that if temperature becomes a greater

issue it may affect satisfaction with the service negatively. Likewise, the location of the final bus stop has a negative correlation between satisfaction and importance, suggesting that those who are mostly satisfied with the location of the bus stop at the end of the journey find this aspect of their journey to be unimportant whilst those who least satisfied with this aspect of their trip, believe this attribute to be highly important.

5.3. Future research and concluding thoughts

The findings of the paper however raise a number of additional questions and issues. Firstly, a question arises as to whether there exists a causal relationship between satisfaction and importance for each of the items, or simply just correlation between the two. That is, it is plausible that once a respondent exceeds some level of satisfaction for a particular trip characteristic, the item may cease to be of importance to them. A similar situation exists within the automobile markets of several countries where through legislation, all vehicles sold must meet a minimum safety standard and hence safety has become less important to consumers in such markets given that all vehicles present are considered to be safe. Likewise, in the case study explored herein, it is possible that once a transit authority has exceeded some satisfaction threshold for a particular item, the item becomes less important to the respondent. Understanding whether such a casual mechanism exists or not is beyond the current study, however the use of repeated sampling over time, where transit authorities implement strategies targeting areas associated with less satisfaction and observing whether the mean importance level for those areas decreases represents one potential way forward. An alternative approach may be to gather information on the threshold levels of both importance and satisfaction and incorporate these somehow into the modelling process.

Secondly, although not addressed herein, the paper has raised similar concerns as others as to whether one should treat the preferences for best and worst questions as symmetrical or not. Clearly, as shown elsewhere, there exist scale differences between the two response types, however whether respondents display similar but opposite preferences for each item is an empirical question (see e.g., Dyachenko et al., 2014; Rose, 2014). Unfortunately, in the current context, assuming at least two mean preference parameters are constrained to be the same for both best and worst, the number of parameter estimates for the model would increase from 129 to 256, making the problem largely intractable given the number of random parameters to be estimated. Future research examining how to address this problem is urgently required given the increasing number of papers appearing within the published literature identifying the existence of this problem in various best worst data sets.

Thirdly, in the current paper, it is worth noting that we have used a forced choice task for both the best and worst questions. Future research should consider how best to incorporate a no choice option, or an equal preference alternative into such questions. How to best ask such questions is not straightforward given that one may be equally satisfied or equally dissatisfied with all the items shown, not to mention believe that all items are similarly equal in terms of importance or unimportance. Alternatively, it is possible that for a given respondent, they are more satisfied (or dissatisfied) with one item shown, but equally satisfied or dissatisfied with the remaining items. A similar pattern is possible for the importance questions. How to first ask such questions, and secondly how to model such data are open questions left to future research.

Aside from the above, the proposed methodology also raise interesting opportunities for transit authorities or regulators interested in measuring the degree of satisfaction for public transportation options in their area of operation. In the current study, the sampling frame consists of all persons who had access to a bus in the Sydney metropolitan region, although we have limited our research to only those who currently use a bus. In a non-academic setting, it might be possible to sample enough respondents drawn from either a specific operator or even from a specific route, hence allowing for the possibility to benchmark down to that level. Even if sample sizes sufficient to allow for operator or route specific models are not feasible, collecting information from respondents about which operator(s) they use and or which routes they currently take may provide sufficient information to allow for differences to be detected across operators and or routes. Given such information, this form of data may either enter the model as covariate information directly allowing for differences to be detected, or one may attempt to look for relationships post estimation by relating such information to outputs such as conditional parameter estimates that can be obtained from the models reported here (see e.g., Train, 2009). Given that the number of possible parameters will increase significantly given the former approach, the use of conditional parameter estimates may represent a simple approach to exploring these issues in the future.

Acknowledgements

The authors would like to thank Anthony Marley for his valuable comments that have improved the paper. Additionally we would like to acknowledge the input of three anonymous reviewers who have also substantively enriched this paper.

References

Abou Zeid, M., 2009. Measuring and Modeling Travel and Activity Well-Being Ph.D. Thesis. Massachusetts Institute of Technology.
Australian Bureau of Statistics, 2011. 2011 Census QuickStats, Retrieved 3rd November 2014.
Auger, P., Devinney, T.M., Louviere, J.J., 2007. Using best–worst scaling methodology to investigate consumer ethical beliefs across countries. J. Bus. Ethics 70 (3), 299–326.
Beirao, G., Cabral, J.A.S., 2007. Understanding attitudes towards public transport and private car: a qualitative study. Transp. Policy 14 (6), 478–489.
Bierlaire, M., 2003. BIOGEME: a free package for the estimation of discrete choice models. In: Proceedings of the 3rd Swiss Transportation Research Conference, Ascona, Switzerland.

Bierlaire, M., Fetiarison, M., 2009. Estimation of discrete choice models: extending BIOGEME. In: 9th Swiss Transport Research Conference, September, Monte Verita, Ascona, Switzerland.
Bouf, D., Hensher, D.A., 2007. The dark side of making transit irresistible: the example of France. Transp. Policy 14 (6), 523–532.
Chou, P., Lu, C., Chang, Y., 2014. Effects of service quality and customer satisfaction on customer loyalty in high-speed rail services in Taiwan. Transport. A 10 (10), 917–945.
Cohen, S.H., Markowitz, P., 2002. Renewing Market Segmentation: Some new tools to correct old problems. In: ESOMAR 2002 Congress Proceedings. ESOMAR, Amsterdam, The Netherlands, pp. 595–612.
Cohen, S.H., Neira, L., 2003. Measuring preference for product benefits across countries: overcoming scale usage bias with Maximum Difference Scaling. In: ESOMAR 2002 Congress Proceedings, Amsterdam, The Netherlands.
Craig, C.S., Douglas, S.P., 2000. International Marketing Research. Wiley, New York.
Dyachenko, T., Walker Reczek, R., Allenby, G.M., 2014. Models of sequential evaluation in best–worst choice tasks. Mark. Sci. 33 (6), 828–848.
de Ona, R., Eboli, L., Mazzulla, G., 2014. Key factors affecting service quality in the Northern Italy: a decision tree approach. Transport 29 (1), 75–83.
de Ona, J., de Ona, R., Eboli, L., Mazzulla, G., 2015. Heterogeneity in perceptions of service quality among groups of railway passengers. Int. J. Sustain. Transport. 9 (8), 612–626.
Dell'Olio, L., Ibeas, A., Cecin, P., 2011. The quality of service desired by public transport users. Transp. Policy 18 (1), 217–227.
Eboli, L., Mazzulla, G., 2008. A stated preference experiment for measuring service quality in public transport. Transport. Plann. Technol. 31 (5), 509–523.
Fishbein, M., Ajzen, I., 1975. Belief, Attitude, Intention and Behaviour: An Introduction to Theory and Research. Addison-Wesley Publishing Company, Reading, Massachusetts.
Finn, A., Louviere, J.J., 1992. Determining the appropriate response to evidence of public concern: the case of food safety. J. Public Policy Mark. 11 (2), 12–25.
Flynn, T.N., Louviere, J.J., Peters, T.J., Coast, J., 2008. Estimating preferences for a dermatology consultation using best–worst scaling: comparison of various methods of analysis. BMC Med. Res. Methodol. 8 (76).
Flynn, T.N., Marley, A.J.J., 2014. Best-worst scaling: theory and methods. In: Hess, S., Daly, A. (Eds.), Handbook of Choice Modelling. Edward Elgar, Cheltenham UK.
Gaymer, S., 2010. Quantifying the impact of attitudes on a shift towards sustainable modes. In: Proceedings of the Australasian Transport Research Forum 2010, Canberra, Australia.
Greene, W.H., 2007. Nlogit Version 4.0 Reference Guide. Econometric Software, New York.
Gilbert, G., Foerster, J.F., 1977. The importance of attitudes in the decision to use mass transit. Transportation 6 (4), 321–332.
Hensher, D.A., Stopher, P., Bullock, P., 2003. Service quality: developing a service quality index in the provision of commercial bus contracts. Transp. Res. Part A 37 (6), 499–517.
Hess, S., Rose, J.M., Hensher, D.A., 2008. Asymmetric preference formation in willingness to pay estimates in discrete choice models. Transp. Res. Part E 44 (5), 847–863.
Hess, S., Train, K.E., Polak, J.W., 2005. On the use of a Modified Latin Hypercube Sampling (MLHS) approach in the estimation of a Mixed Logit model for vehicle choice. Transp. Res. Part B 40 (2), 147–163.
Lee, J.A., Soutar, G.N., Louviere, J.J., 2007. Measuring values using best–worst scaling: the LOV example. Psychol. Market. 24 (12), 1043–1058.
Lee, J.A., Soutar, G., Louviere, J.J., 2008. The best–worst scaling approach: an alternative to Schwartz's values survey. J. Pers. Assess. 90 (4), 335–347.
Marley, A.A.J., 2010. The best–worst method for the study of preferences: theory and application. In: Frensch, P.A., Schwarzer, R. (Eds.), Cognition and Neuropsychology: International Perspectives on Psychological, vol. 1. Psychology Press, Hove, pp. 147–157.
Marley, A.A.J., Flynn, T.N., 2015. Best worst scaling: theory and practice. In: Wright, J.D. (Ed.), International Encyclopedia of the Social and Behavioral Sciences, second ed. Elsevier, London UK.
Marley, A.A.J., Flynn, T.N., Louviere, J.J., 2008. Probabilistic models of set-dependent and attribute-level best–worst choice. J. Math. Psychol. 52 (5), 281–296.
Marley, A.A.J., Louviere, J.J., 2005. Some probabilistic models of best, worst, and best–worst choices. J. Math. Psychol. 49 (6), 464–480.
Mielby, I.H., Edelenbos, M., Thybo, A., 2012. A comparison of rating, best–worst scaling, and adolescents' real choices of snacks. Food Qual. Prefer. 25 (2), 140–147.
Mowen, J.C., 1993. Consumer Behaviour, third ed. Macmillian Publishing Company, New York, pp. 151–182, Chapter 7.
Paine, F.T., Nash, A.N., Hille, S.J., Allen, G., 1969. Consumer attitudes toward auto versus public transport alternatives. J. Appl. Psychol. 53 (6), 472–480.
Paulhus, D.L., 1991. Measurement and control of response bias. In: Robinson, J.P., Shaver, P.R., Wrightsman, L.S. (Eds.), Measures of Personality and Social Psychological Attitudes. Academic Press, New York, pp. 17–59.
Recker, W.W., Stevens, R.F., 1976. Attitudinal models of modal choice: the multinomial case for selected non-work trips. Transportation 5 (4), 355–375.
Roman, C., Martin, J.C., Espino, R., 2014. Using stated preferences to analyse the service quality of public transport. Int. J. Sustain. Transport. 8 (1), 28–46.
Rose, J.M., 2014. Interpreting discrete choice models based on Best–Worst data: a matter of framing. In: Transportation Research Board, Annual Meeting, January 12–16, Washington D.C.
RSG, 2013. Bus Rapid Transit Focus Groups and MaxDiff Surveys. <http://www.rsginc.com/node/109> (accessed 29/02/15).
Scarpa, R., Notaro, S., Louviere, J.J., Raffaelli, R., 2011. Exploring scale effects of best/worst rank ordered choice data to estimate benefits of tourism in alpine grazing commons. Am. J. Agric. Econ. 93 (3), 813–828.
Spitz, G.M., Greene, E.R., Adler, T.J., Dallison, R., 2007. Qualitative and quantitative approaches for studying transit station. In: Proceedings of the 86th Annual Meeting of the Transportation Research Board, Washington, United States.
Stathopoulos, A., Marcucci, E., 2014. De Gustibus Disputandum Est: Non-Linearity in Public Transportation Service Quality Evaluation. Int. J. Sustain. Transport. 8 (1), 47–68.
Steenkamp, J.B., Baumgartner, H., 1998. Assessing measurement invariance in cross-national consumer research. J. Consumer Res. 25 (1), 78–90.
SMTC, 2011. The I-81 Challenge: Spring 2011 Questionnaire Summary. Syracuse Metropolitan Transportation Council. <http://thei81challenge.org/cm/ResourceFiles/resources/Questionnaire%20Summary _FINAL.pdf> (accessed 01.11.13).
TCRP, 2008. Understanding how individuals make travel and location decisions: implications for public transportation. In: Transit Cooperative Research Program Report 123, Washington, United States.
Thompson, K., Schofield, P., 2007. An investigation of the relationship between public transport performance and destination satisfaction. J. Transp. Geogr. 15 (2), 136–144.
Train, K., 2009. Discrete Choice Methods with Simulation, second ed. Cambridge University Press.
Tukey, J., 1949. Comparing individual means in the analysis of variance. Biometrics 5 (2), 99–114.
Zhang, K., Zhou, K., Zhang, F., 2014. Evaluating bus transit performance of Chinese Cities: developing an overall bus comfort model. Transp. Res. Part A 69, 105–112.