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