Transport Reviews: A Transnational Transdisciplinary Journal
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/ttrv20

Passengers' Perceived Service Quality of City Buses in Taipei: Scale Development and Measurement

Kai-Chieh Hu [a] & William Jen [b]

[a] Department of Shipping and Logistics Management , Kainan University , Taiwan

[b] Department of Transportation Technology and Management , National Chiao Tung University , Taiwan

Published online: 23 Feb 2007.

To cite this article: Kai-Chieh Hu & William Jen (2006) Passengers' Perceived Service Quality of City Buses in Taipei: Scale Development and Measurement, Transport Reviews: A Transnational Transdisciplinary Journal, 26:5, 645-662

To link to this article: http://dx.doi.org/10.1080/01441640600679482

Transport Reviews, Vol. 26, No. 5, 645–662, September 2006

Passengers' Perceived Service Quality of City Buses in Taipei: Scale Development and Measurement

KAI-CHIEH HU* and WILLIAM JEN**

**Department of Shipping and Logistics Management, Kainan University, Taiwan **Department of Transportation Technology and Management, National Chiao Tung University, Taiwan*

(Received 14 December 2005; revised 14 February 2006; accepted 9 March 2006)

ABSTRACT *To understand fully passengers' perceptions and expectations of the bus service quality in Taipei, business managers and governmental agencies must seek a proper scale that can reflect passengers' opinions accurately. This study develops and tests a service quality scale designed for a city bus transit system in Taipei. Churchill's paradigm and a focus group interview were combined into a multistage scale development procedure. Based on the procedure, Taipei city buses were selected as the example, for which a service quality scale was developed. The final scale contains four dimensions and 20 items. These four dimensions are 'interaction with passengers', 'tangible service equipment', 'convenience of service' and 'operating management support'. Finally, the results of scale development and the managerial applications of the service quality scale for the city transit system are discussed.*

Introduction

Public transportation service is important in determining quality of life (Cunningham *et al.*, 2000). With more than 2.6 million people, Taipei is a dense metropolis with congested traffic in an area of approximately 272 km^2. Although the Taipei mass rapid transit system has been in operation since 1996, most people who live or work there still must take a bus to reach or transfer to the Mass Rapid Transit (MRT) to reach their trip destination. Therefore, the city bus system remains one of the main mass transit systems in the Taipei metropolitan area. The Taipei city bus operators include one government agency and 14 private companies (five of these companies had fewer than three routes during the investigating period). They operate a total of 264 routes, have more than 2000 stops and over 1.7 million passengers use the bus system every day.

The quality of the bus service is one of the key factors that affects the willingness of the public to ride a bus or drive their cars during the rush hours. The need to maintain market share and increase profitability is the main driving force to improve quality in a mass transit service (Pullen, 1993). Therefore, it is

Correspondence Address: Kai-Chieh Hu, Department of Shipping and Logistics Management, Kainan University, No. 1 Kainan Road, LuChu, Taoyuan County 338, Taiwan. Email: hkchieh@mail.knu.edu.tw

0144-1647 print/1464-5327 online/06/050645-18 © 2006 Taylor & Francis
DOI: 10.1080/01441640600679482

very important to measure the perception and expectation that passengers have of the service quality. To govern the level of bus service, the Taipei City Government (TCG) has been handling the 'Evaluation of Taipei City Bus' Operation and Service' since 1993. This evaluation, which was designed in terms of supervision, contains four dimensions: facilities and services of the bus station; vehicle safety and security of transport; the quality of passenger service and management of the bus drivers; and the administration and management of the company. The TCG data survey is being conducted by an independent consultant, but they do not investigate the experience and perceptions of the actual passengers. Thus, they are unable to determine whether the results agree with the perceptions and expectations that passengers have of the service quality. Earlier researches of mass transit service were focused on the measurement of productivity and performance (e.g. Botzow, 1974; Alter, 1976; Bakker, 1976; Fielding and Anderson, 1983; Metropolitan Transit Authority of Harris County, 1984; Fielding *et al.*, 1985; Carter and Lomax, 1992; Fielding, 1992; Pullen, 1993; Wipper, 1993; Hensher and Daniels, 1995), but little attention was paid to service quality. Several of the later studies refocused on the measurement of service quality (Hensher *et al.*, 2003; Buchanan, 2005; Transport for London, 2006). Hensher *et al.* (2003) establish a way to measure and calculate an overall service quality index (SQI), which includes 13 attributes, such as bus travel time, bus fare, time walking to the bus stop, seat availability, information, driver attitude, etc. However, non-academic monitoring of passengers' satisfaction on city buses is already ongoing in several cities around the world. The London Buses quality of service is regularly being observed and it contains 11 indices, such as safety, crowding, reliability, cleanliness, staff behaviour, etc. (Transport for London, 2006). The Department for Transport of England (2005) published a report on the transport statistics of bus and light rail service in England. The bulletin surveyed bus passenger satisfaction using 17 measures, which were aggregated into five composite indicators: safety and security; bus stop/shelter condition; condition of the bus; journey speed; and staff service/comfort. The result showed that the average score for overall satisfaction was 80 out of 100, down one point from the previous quarter. In addition, there are periodical satisfaction surveys for bus passengers in other areas or countries, such as in Scotland, UK (Buchanan, 2005), Europe: the MARETOPE project (TIS.PT, Consultores em Transportes, Inovação e Sistemas, S.A., 2003), the Netherlands (AVV Transport Research Centre, 2004) and Singapore (Public Transport Council of Singapore, 2005). Although most of them adopted similar indicators and did a very thorough investigation, no further research was ever carried out to confirm whether the attributes were actually related to the concerns and desires of the passengers.

Service quality management cannot be successful without accurately assessing customers' quality perceptions (Cunningham *et al.*, 2000). The present authors think the apposite way to evaluate bus passengers' service quality is from the viewpoint of marketing management. Parasuraman *et al.* (1988) devise ten basic dimensions regarding service quality and design a questionnaire to conduct investigations. These dimensions are tangibles, reliability, responsibility, competence, courtesy, credibility, security, access, communication, and understanding/knowing the customer. Based on which, Parasuraman *et al.* conduct various experiments within various industries and choose five that are best suited to measure service quality in these industries. Thus, they propose a service quality scale (SERVQUAL) by simplifying the original scale that is comprised of 22 items

and five cognitive dimensions (tangibility, reliability, responsiveness, assurance and empathy). Parasuraman *et al.* believe that SERVQUAL was practical on all types of service businesses. Following some research (Carman, 1990; Finn and Lamb, 1991; Babakus and Boller, 1992; Cronin and Taylor, 1992; Triplett *et al.*, 1994), however, these five basic dimensions and the 22 items were determined to be inappropriate for some service businesses. The current methods of measuring general service quality cannot be applied directly to assess the service quality of mass transit. To develop an exhaustive measurement, a systematic procedure should be available to develop a service quality scale for mass transit.

The present study develops and assesses a service quality scale for Taipei's bus service. Although the previous studies have measured the service quality perception of bus service, this study is the trailblazer that applies the SERVQUAL method and Churchill's (1979) paradigm to develop a service quality scale of the city bus. The present paper provides a useful example of a case study in developing a suitably condensed measurement scale of bus service. To improve the validity of the scale, Churchill's paradigm and a focus group interview were combined into a multistage procedure for scale development, which is an iterative process of quantitative statistic analysis and qualitative approach. The procedure contains ten steps in four stages: make an elementary exploration of the scale; consult the viewpoints of customers; analyse the modified scale; and purify the items and dimensions. The focus group interview is conducted in the second stage for two reasons. First, since each researcher had a limited understanding of the actual system, the group's opinions clarified several issues. Second, the focus group interview clarified and predicted problems that might have occurred in the formal investigation. The inarticulate words and inadequate statements were removed to prevent the low reliability and validity.

Notably, Taipei city bus was selected as the example and on which a service quality scale was created. The paper is as follows. Initially, the procedure of the service quality scale's development is explained. The improvement process of the scale development and dimensions' changes is then reported. Finally, the study results from these procedures and applications for managers are discussed.

Scale Development of Service Quality

DeVellis (1991) indicates that a measurement tool, or scale, should include questions to reflect theoretical variants, which the observations may not reveal. It is often much more efficient to measure the dimensions based on scale when a concrete indicator is not available. Churchill (1979) proposes a framework that involves an iterative process to develop marketing measures. In the first stage, a researcher must specify the domain of the construct and generate a sample of items to be measured. The following steps are data collection and purification, which are based on factor analysis and on the calculation of Cronbach's alpha (Cronbach and Meehl, 1955). An assessment of reliability and validity of the construct follows the second stage of data collection (Smith, 1999).

Due to the features of general service businesses, intangibility, inseparability, variability and perishability, service quality is more difficult to measure than is hardware quality (Parasuraman *et al.* 1985). Parasuraman *et al.*'s SERVQUAL and its related research are considered the most comprehensive in service quality. Their basic premise is that service quality should be the gap between customers' expectations and perceptions. Hence, if the perceived service is less than

expected, then service quality is poor. In contrast, if the opposite is true, service quality is superior. Parasuraman *et al.* discuss ten dimensions of passengers' perceived service quality and investigate each one with experiments and tests (Parasuraman *et al.*, 1988, 1991; Berry *et al.*, 1990; Zeithaml *et al.*, 1990). Their scale-developing procedures, which has 11 steps and two stages, closely parallels procedures recommended in Churchill's paradigm when developing a general scale from customers' viewpoint (Parasuraman *et al.*, 1988). Although SERVQUAL serves as the scale for varying industries, the resulting dimensions of service quality from other researches were not the same (Bitner, 1990; Cronin and Taylor, 1992; Boulding *et al.*, 1993; Brown *et al.*, 1993). Carman (1990) recommends that to obtain reliable results, questionnaires should be designed based on the original ten dimensions.

In addition to the work of Parasuraman *et al.*, some investigators have investigated the steps and procedures for scale development, and some have focused on statistical analysis (Bienstock *et al.*, 1997; Mentzer *et al.*, 1999; Shemwell and Yavas, 1999). Smith (1999) thinks that when adopting Churchill's paradigm, there are shortcomings in specifying the domain of the construct and, as a result, important criteria may be deleted. McCormack (1994) contends that although SERVQUAL neglects qualitative aspects, it should be applied to quantify service quality. Furthermore, to evaluate service quality, he recommends considering the qualitative method. For example, focus group interview can be conducted. Fowler (1995) thinks that the focus group interview could improve a questionnaire in two ways. First, the related hypothesis regarding the investigation context can be examined. Second, it aids in evaluating the expressions in the questionnaires and the hypothesis regarding the language used or cognitive assumptions. Cunningham *et al.* (2000) provide a methodological framework that can be employed to determine service quality. Their empirical results confirm that both quantitative and qualitative research techniques should be used to understand customer perceptions and service quality evaluations more thoroughly (Morgan, 1988; Krueger and Casey, 2000).

Although other scales have been developed for specific purposes, such as Dabholkar *et al.*'s (1996) retailer scale, Bouman and Van der Wiele's (1992) car service scale, Mentzer *et al.*'s (1999) logistics scale, and Shemwell and Yavas's (1999) hotel scale, these studies fail to recommend how to incorporate both the qualitative and quantitative in these procedures. In addition, chiefly due to the variety of its customers, the quality of mass transit service is difficult to gauge. Restated, these investigations cannot adequately represent differing customer opinions regarding service.

Scale Development Procedure

This study works with a multiple methodologies procedure derived from Churchill's paradigm to obtain a reliable and valid scale. The complete procedure has four stages and ten steps. Figure 1 illustrates the procedure and development steps.

The purpose of stage 1 is to make an elementary development of the domain of service quality, and then reduce the number of items of a scale. Customers' activities must first be clearly understood to ensure that items are designed properly and scale dimensions verified (step 1). Then, based on Zeithaml *et al.*'s (1990) ten dimensions, the questions are confirmed to include each step in the activity flow

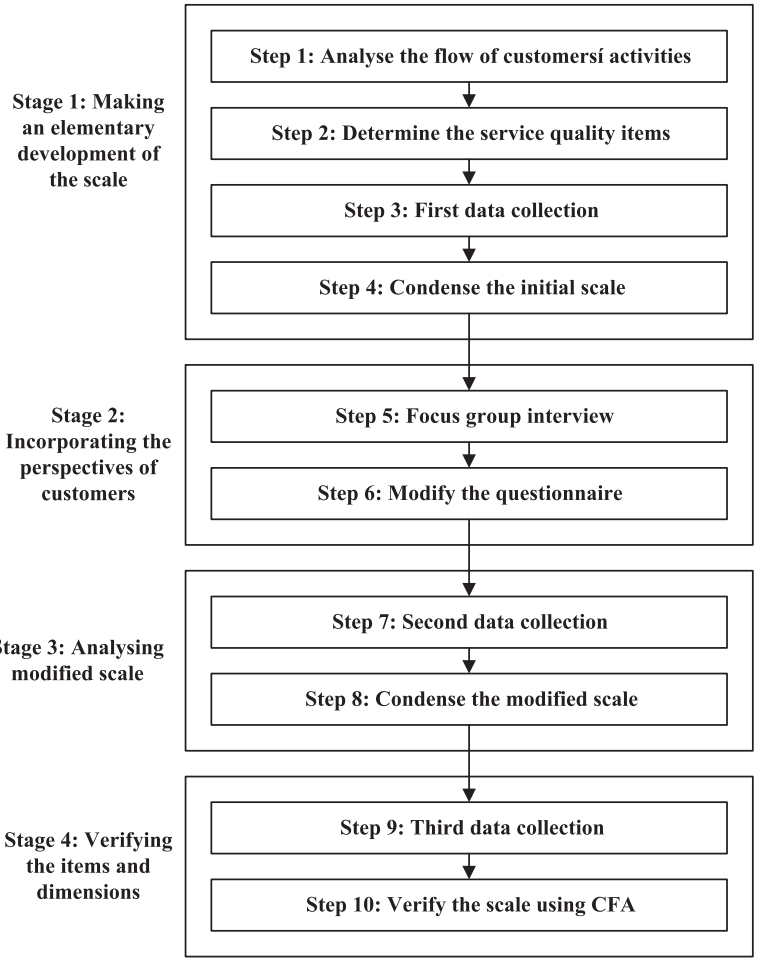

Figure 1. Procedure for developing service quality scale

(step 2). After the questionnaire is completed, a pilot study is conducted (step 3). Customers' expectations and perceptions of service can then be determined. The gap between the perception and expectation reveals the service quality of each item. The sample size is required to exceed 200, according to Hatcher (1998). After the data are collected, an iterative reliability analysis and explanatory factor analysis are applied to the items (step 4). Inappropriate questions are eliminated and the dimensions of service quality are extracted and named appropriately. Figure 2 presents the statistical analytical flow of condensing questionnaires.

The purpose of stage 2 is to incorporate the perspective of customers. The qualitative method is applied in steps 5 and 6, in which the focus group is interviewed and the scale modified. Fowler (1995) thinks that a focus group interview could improve a questionnaire in two ways. First, since each researcher had a limited understanding of the actual system, the group's opinions clarified several issues. Second, the focus group interview clarified and predicted problems that might have occurred in the formal investigation. The participants initially discuss the first scale and related topics. The outcome of the focus group interview justifies the deletion of a basis for deleting inadequate items and the combination of questions.

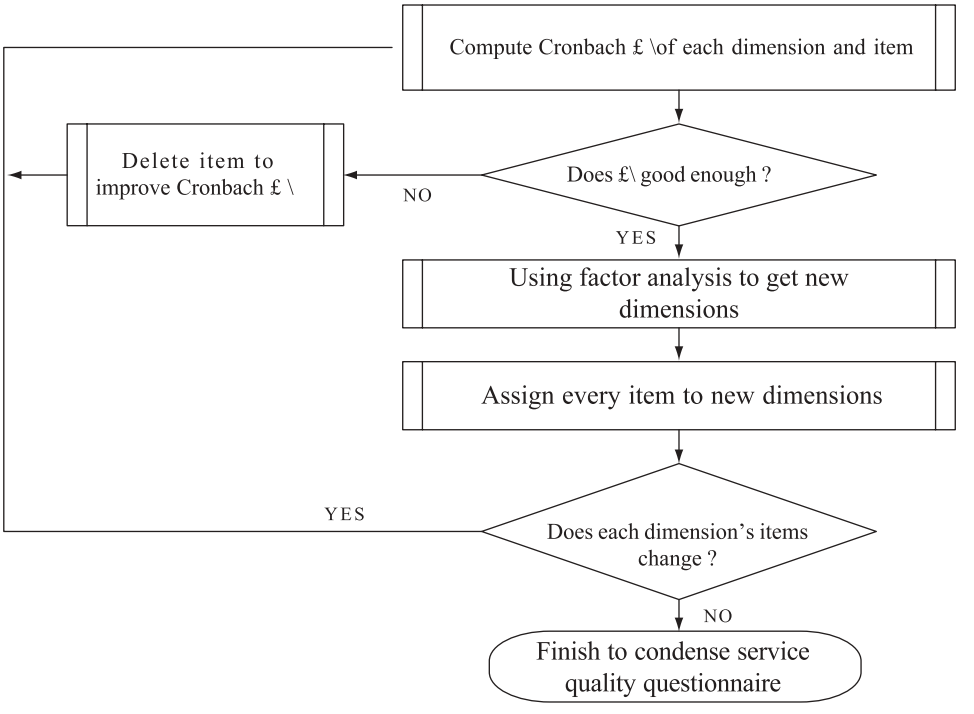

Figure 2. Process of questionnaire condensing

It also indicates if items that the focus group determines are significant should be added. The statements and types of response are also modified in developing a new questionnaire (step 6). Thus, in stage 3, new data must be collected to simplify and analyse the service quality scale again. The same statistical process as in step 4 is executed again to verify the dimensions of the modified scale (steps 7 and 8). Service dimensions are then determined and a second simplified scale is presented. Finally, in stage 4, confirmatory factor analysis (CFA) is applied to the data sets collected in step 9 to verify the items and dimensions of the scale. The composite reliability and convergent validity are also analysed to determine the stability of the scale (step 10).

Case of Taipei City Bus Service

This study considers the bus service in Taipei as an empirical case to provide a useful example of a case study in developing a suitably condensed measurement scale. The results from the steps are summarized on Table 1 and the detail mentioned is as follows.

Stage 1: Making an Elementary Development of the Scale

Step 1: Analyse the flow of customers' activities. Initially, the relationship between the flow of passengers' activities and the bus service system was analysed using a simplified service blueprint method (Figure 3). The flow of activities involved in taking a bus can be divided into many steps. The passenger goes to the bus stop from a start point. When the bus stops, the passenger gets on it. The bus then

Table 1. Summarized results of each step

Step 3			Step 4		Step 6	Step 7	Step 8		Step 9
Dimensions	Items	Alpha	Dimensions	Items	Items	Alpha	Dimensions	Items	Alpha
Tangibles	16	0.86	Interaction with Passengers	12	6	0.92	Interaction with passenger	6	0.82
Reliability	6	0.76							
Responsiveness	5	0.78	Tangible service	7	6	0.86	Tangible service	6	0.84
Competence	3	0.68	Equipment				Equipment		
Courtesy	2	0.76	Operating management support	4	4	0.89	Convenience of service	5	0.81
Credibility	2	0.40					Operating management support	3	0.90
Security	4	0.80	Convenience of service	4	4	0.84			
Access	4	0.68	Offering correct information	3	3	0.89			
Communication	4	0.85							
Understanding the customer	4	0.67							
Total scale	50	0.95		30	23	0.96		20	0.95
Cumulative explained variance (%)		57.21				70.93			–[a]

The analysis method on the third data sets was confirmatory factor analysis (CFA) with no cumulative explained variance value.

travels along its route. When the bus arrives at the passenger's destination, the passenger alights. Finally, he/she walks to his/her destination. Notably, the entire activity flow occurs outside, and is easily disturbed by urban traffic, traffic signs, road paving and even the weather. Several factors affect a passenger's willingness to take a bus. However, this work focuses primarily on improving the quality of service offered by bus service providers.

Step 2: Determine the service quality items. Carman (1990) suggests that researchers should design different questionnaires for different industries, with ten dimensions. Based on these dimensions and the activity flow, 64 questions were initially posed. Following limited sample testing, the collection of opinions during testing, and a discussion among a study team, 50 items were selected. The questionnaire had two parts—one addressing expectation and the other perception. In the first part, passengers were asked which level of service they thought the bus industry should provided. The second part asked how the passengers perceived the service. All answers were given on a Likert seven-point scale, from 'agree very strongly' to 'disagree very strongly'. A space was included to enable people to note their opinions regarding the questionnaire or ways to improve the quality of service.

Step 3: First data collection. The questionnaire was distributed as a mail survey. The respondents were 'volunteers' recruited by the Bureau of Transportation of TCG. TCG had recruited 1000 bus passengers as 'volunteers' to answer a modal questionnaire designed by the Bureau of Transportation. These volunteers, who were all enthusiastic, had much experience of riding on buses. Consequently, 500 subjects were selected at random from these volunteers in the first stage. Questionnaires were then sent by mail and the subjects asked to complete them in 3 weeks.

Of the 500 mailed questionnaires, 244 (48.8%) valid questionnaires were received. The ages of the respondents ranged between 16 and 35 years; 74.6% were female. Students and office workers represented 25.0 and 53.3% of the respondents, respectively. Most respondents graduated from senior high school (85.7%).

Step 4: Condense the initial scale. Cronbach's alpha method was used to analyse reliability. Column 3 of Table 1 indicates that most alpha values were exceeded by 0.6, except 'Credibility'. The questionnaire condensing process was then employed. The Bartlett value of the data set was 3587.636 and $p = 0$. Thus, factor analysis could be applied. The Kaiser–Meyer–Olkin (KMO) value was 0.914, implying that the data sets are random and appropriate. EFA was then performed to extract the dimensions of service quality. According to Kaiser and Rice (1974), a common factor with an eigenvalue > 1 should be reserved. As no obvious factor loading dominated these main factors, orthogonal rotation was performed in varimax rotation to determine the factor loading (Kaiser and Rice, 1974). After condensation and iterative factor analysis, a questionnaire with 30 items and five dimensions was produced, with a cumulative explained variance of 57.21%. Accordingly, the five dimensions were identified and named, as shown in column 4 of Table 1.

Stage 2: Incorporating the Perspectives of Customers

Step 5: Focus group interview. The focus group interview aimed to capture distinct service experiences, not representative samples of experience (Stewart, 1990). Thus, people who had extensive experience of taking buses and were enthusiastic

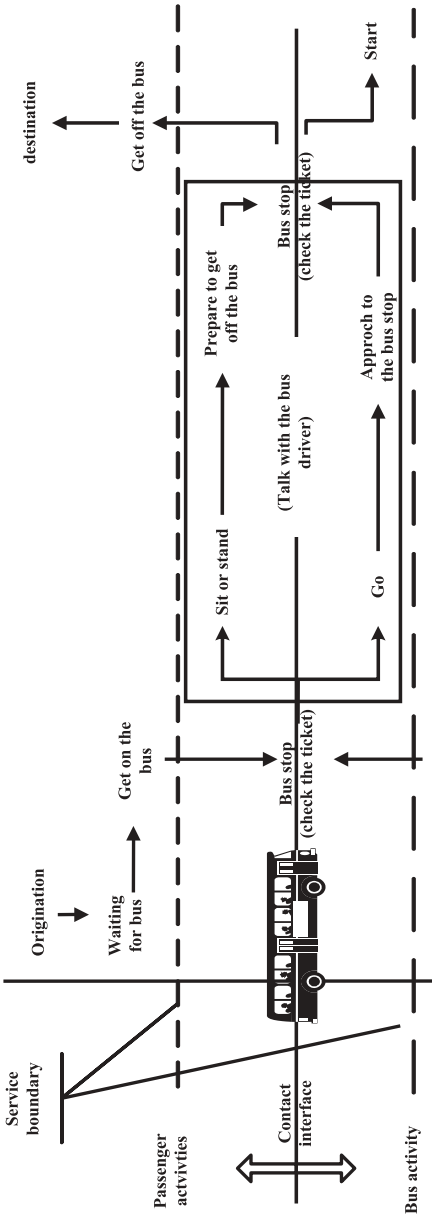

Figure 3. Passengers' activity flow: a simplified service blueprint

about offering suggestions were selected to yield a broad range of information from the group. Some respondents, who in the first stage provided several opinions in response to the questionnaire, were selected as members of the focus group. Sixty-one invitations were mailed and 20 volunteers participated in the discussion. The group discussed topics ranging from the service items in the questionnaire to the simplified scale.

The focus group performed two main activities: answering the questionnaire and discussing service items that were considered to be significant. Each participant responded to the questionnaire. The questionnaire items they had evaluated were then discussed. The modification lies in the deletion of some inadequate items, the combination of some similar questions, the revision of some statements and the inclusion of the services that the focus group thinks very important. There are some items revised in the statements to show integrity and representativeness. The revision helps one avoid the chance that the passengers do not know how to answer the questions because of confusion.

Step 6: Modify the questionnaire. Further to the focus group interview, the scale was modified. Modifications included deleting inadequate items, combining similar questions, and revising and including services items, all of which the focus group considered vital. For example, a question regarding ambient light was dropped because of its limited importance. Items were revised to indicate integrity and representation. Moreover, the revisions reduced confusion. After combinations and deletions, 23 items were left and they formed the new scale.

One of the modifications was to have the tolerance range in the expected service that Parasuraman *et al.* (1985, 1988, 1991) came up with in their studies. The tolerance range is the distance between the desired and the tolerable. Service quality should be the distance between the perceived and the tolerance range. In this way, the outcome will be more reasonable and more accurate. The modification also explores the necessity to analyse the tolerance range further. In addition, Parasuraman *et al.* (1985, 1988, 1991) suggested that answers provided should be in the format of three columns from the previously used Likert seven-point scale since the expectation section is divided into two parts—the desired and the tolerable. However, the discussion in the focus group indicates that the format of the seven-point scale is not enough to answer three questions. To overcome the problem, the way to give answers is changed into the ten points, and the responders must answer three parts: the desired service, the tolerable service and the perceived service.

Stage 3: Analysing the Modified Sale

Step 7: Second data collection. The modified questionnaire was then distributed to passengers at bus stops, i.e. passengers received a copy when waiting for or disembarking from a bus. Of the 800 distributed, 292 (36.5%) valid questionnaires were returned via the mail. Primarily, the age range of passengers was 16–25 years (49%). However, 25.7% of respondents were 26–35 years of age. The male-to-female ratio was approximately 4:6. The majority of respondents were either students (42.5%) or office employees (43.0%). The samples exhibited similar demographic characteristics to those involved in other research and the first survey in this study.

Step 8: Condense the modified scale. As the above previous analysis, a reliability analysis was performed in each dimension and on the entire scale. Column 7 of

Table 1, which shows Cronbach's alpha values, indicates a high reliability in excess of 0.8. The entire scale had an alpha value of 0.958, which meant this modified scale was more reliable than the previous one. Following the reliability analysis, the scale was simplified and a dimensional analysis performed. Stable dimensions, associated with 20 items, were confirmed after iterative analysis. Four final dimensions, explaining 70.93% of the variance, were selected. The dimensions were as follows:

- Interaction with passengers. Passengers feel respected and cared for when interacting with service providers. The bus companies' responses to the passengers' problems were friendly and reasonable.
- Tangible service equipment. The bus companies offer the passengers a comfortable facility and hardware.
- Convenience of service. Passengers have easy access to the bus service. They are offered sufficient bus information and a comfortable waiting facility.
- Operational management support. Bus companies' support satisfies passengers' demand, including the bus schedule, service period, dispatch of attendants, etc.

Column 8 of Table 1 lists each dimension; Column 9 presents the number of items for each dimension that resulted from this step.

Stage 4: Verifying the Items and Dimensions

Step 9: Third data collection. New data were collected from passengers at bus stops to determine the reliability and validity of the scale obtained in step 8. Of the 500 distributed questionnaires, 235 (31.3%) were returned. Most of the respondents were aged 16–35 years (63.4%); 61.5% were females. The largest groups of respondents were students (45.5%) and office employees (37.8%).

Step 10: Verify the scale using CFA. Again, Cronbach's alpha values were computed to assess the reliability of each dimension and the scale. The results in column 10 of Table 1 show that all Cronbach's alpha values were high and exceeded 0.8. Thus, this modified scale is reliable. Although the statistical analysis indicated that the scale would have a higher alpha value, CFA was performed in this stage to make a further confirmation. By using CFA, $x^2 = 412.38$ ($p < 0.0001$), and the degrees of freedom (d.f.) were 164. The common level of the x^2/d.f. ratio was less than 5, though below 3 was better (Joreskog and Sorbom, 1993). The x^2/d.f. ratio in the present study was 2.51, which indicated good fitness. Furthermore, the model fit indicators were for CFI (Bentlers comparative fit index) = 0.855, GFI (goodness of fit index) = 0.838, AGFI (adjusted for degrees of freedom) = 0.792 and RMR (root mean square residual) = 0.08, indicating that the dimensions of the final scale fitted the data.

The reliability and validity of the scale were also assessed by composite reliability and factor loadings, as shown in Table 2. The composite reliability of each construct exceeded 0.7 in this study, and was thus satisfactory (Hatcher, 1998). On the other hand, all *t*-values for all items ranged from 3.848 to 16.505, implying that all factor loadings were significant ($p < 0.001$). These results provide evidence of convergent validity of all indicators that effectively measure a single construct (Anderson and Gerbing, 1988). According to these results, the dimensions of the final scale were stable.

The foregoing analysis indicated that the constructed items that were derived from our procedure for developing a scale were highly reliable. These items can thus accurately reflect and measure the service quality.

Table 2. Confirmatory factor analysis (CFA) results: the third data collection

Variable	Standardized factor loadings	*t*	Composite reliability
Interaction with passengers:			0.774
Drivers appreciate the safety of passengers when they get on/off the bus	0.681	10.456*	
Drivers are polite and friendly when communicating with passengers	0.652	9.884*	
Drivers drive buses smoothly, and their road-craft is fine	0.659	10.021*	
Drivers drive on the right route and never fail to stop when passengers want to get on	0.281	3.848*	
Company deals with accidents quickly and reasonably	0.592	8.775*	
Company deals with passengers' opinions and complaints sufficiently	0.717	11.159*	
Tangible service equipment:			0.794
Bus companies provide safe and brand new buses	0.699	10.909*	
Vehicles are clean inside	0.679	10.492*	
Noise on the car is not too loud	0.592	8.851*	
Equipment in the bus satisfies passengers' needs	0.642	9.782*	
Air-conditioning is very comfortable	0.683	10.571*	
Stop's layout is fine	0.477	6.872*	
Convenience of service:			0.78
Places of bus stops are proper and convenient for taking a bus	0.609	9.095*	
Transhipping on the network is convenient	0.655	9.936*	
Information about bus routes is marked clearly	0.657	9.975*	
Company will have notification on the buses in short time when the routes and bus schedule are changed	0.602	8.963*	
Company will correct the information at stops in the short time when the routes and bus schedule are changed	0.697	10.761*	
Operating management support:			0.891
I do not have to worry that there is no bus	0.864	15.420*	
I usually wait for a bus longer than the scheduled headway	0.903	16.505*	
Company dispatches buses according to the schedule	0.798	13.731*	

*Significant value ($p < 0.001$).

Measurement Results of Service Quality

Following the reliability and validity analysis, this service quality scale was employed to measure the performance of the city bus service correctly. The scores of each item and dimension from the final sample were evaluated. Notably, a higher score represented superior service quality. Furthermore, a paired-sample *t*-test was employed to test the significance of differences between the tolerable and the perceived service. Table 3 lists the empirical results of all items and dimensions. Perceived service had a significant difference from tolerable service, which meant that the current bus service satisfied the passengers' lowest expectations.

Among the four dimensions, the first 'interaction with passengers', had a relatively lowest service quality score (0.75), which indicated a poor appreciation of the interaction between staff and passengers. The relatively worse items were managing opinions and complaints as well as drivers' driving ability. The study result is significant since previous managers had minimized operating costs by

Table 3. Assessment of service quality scale across the final sample

	Desired service		Tolerable service		Perceived service		Service quality[a]		Significance
Dimensions and items	Mean	SD	Mean	SD	Mean	SD	Score	*t*	($\alpha = 0.05$)
Interaction with passengers:	9.38	–	5.56	–	6.31	–	0.75	–	–
Q1	9.62	0.76	5.85	1.8	6.6	1.78	0.75	5.9	0
Q2	9.13	1.09	5.23	1.71	6.16	1.65	0.92	8.25	0
Q3	9.45	0.93	5.71	1.76	6.28	1.75	0.57	4.55	0
Q4	9.46	1.02	5.76	1.78	6.77	1.79	1	8.1	0
Q5	9.36	0.97	5.49	1.83	6.28	1.62	0.79	6.55	0
Q6	9.23	1.1	5.29	1.76	5.78	1.78	0.5	3.77	0
Tangible service equipment:	9.21	–	5.21	–	6.54	–	1.33	–	–
Q7	9.73	0.63	5.69	1.52	6.9	1.41	1.21	12.52	0
Q8	9.31	0.92	5.38	1.48	6.67	1.53	1.28	11.76	0
Q9	9.01	1.1	5.01	1.53	6.2	1.5	1.19	11.35	0
Q10	8.87	1.19	4.81	1.54	6.48	1.51	1.67	17.84	0
Q11	9.36	0.94	5.38	1.68	6.73	1.6	1.35	12.2	0
Q12	9	1.19	4.96	1.61	6.25	1.61	1.3	11.23	0
Convenience of service:	9.37	–	5.51	–	6.74	–	1.23	–	–
Q13	9.3	0.94	5.35	1.5	6.78	1.38	1.43	15.4	0
Q14	9.22	1.08	5.33	1.64	6.81	1.46	1.48	15.24	0
Q15	9.47	0.87	5.67	1.72	6.97	1.58	1.3	11.5	0
Q16	9.38	0.96	5.6	1.78	6.66	1.79	1.06	9.8	0
Q17	9.46	0.91	5.62	1.75	6.48	1.83	0.86	6.82	0
Operating management support:	9.49	–	5.62	–	6.59	–	0.97	–	–
Q18	9.5	0.86	5.61	1.68	6.78	1.65	1.17	9.84	0
Q19	9.5	0.84	5.66	1.77	6.51	1.74	0.85	6.45	0
Q20	9.47	0.88	5.6	1.79	6.47	1.72	0.87	6.48	0

[a]Service quality = perceived service – tolerable service.

redesigning bus routes and schedules. That is, they ignored that customer perception and service quality were decreasing.

The second dimension, 'tangible service equipment', obtained the relatively highest score (1.33), which confirmed that passengers received comfortable facilities and hardware. Passengers were pleased with most tangible equipment; loud cars and vehicles with poor exteriors were disliked.

The third dimension, 'convenience of service', had a score of 1.23 and proved that the bus service is accessible. In particular, the bus network that increased convenience had relatively the highest score. However, it received the lowest score in updating information at a stop when the route or bus schedule was modified.

Finally, the dimension 'operating management support' obtained a relatively lower service quality score of 0.97. The chief problem was that passengers were uncertain about reliability and, thus, normally had a long waiting. Hence, waiting time decreased passengers' perceived service quality in this study.

Tolerance Range Analysis

One particular application of this study result was to analyse passengers' tolerance range of each service dimension. The tolerance range is the gap between the tolerable and the desirable service. Figure 4 illustrates the study results of the tolerance range analysis that was based on the data shown in Table 3. The top and bottom short lines represent the value of the desired and the tolerable service, respectively. The little square between the two lines represents the value of the perceived service. The tolerance range analysis assists in the clarification of the improvement priority of the service dimensions. It demonstrates that the perceived service in the 'tangible service equipment' and 'operating management support' were almost identical (Figure 4). However, in the dimension of management support, the gap between the perceived and the tolerable was much smaller than expected. In contrast, in 'tangible service equipment', the variation between the perceived/the tolerable and the perceived/the expected was not as obvious. This difference between the two dimensions proves that management support should take priority over the tangibility in improvement.

Discussion

The study discussed dimensions and scale to measure the service quality of mass transit. Only when the measurement tool is exhaustive and stable could it be applied for further analysis and evaluation. Following completion, the present scale can be used to evaluate bus service quality.

Unlike previous scales, this study attempted to modify Parasuraman *et al.*'s stages and construct a more solid and systematic service scale for the public transportation industry. This study involved four stages. A focus group interview is conducted in the second stage to incorporate the opinions of customers, while CFA is employed in the final stage to assess the stability of the dimensional structure of the final scale. The final scale and dimensions are shown to be highly reliable and valid. Thus, this procedure for developing a scale is proven to be workable and to yield reasonable results.

Table 1 also presents the numbers of dimensions, items and Cronbach's alpha. The ten dimensions are condensed. Although initially, there were five, only four dimensions were included in the final draft. The original 64 items were initially

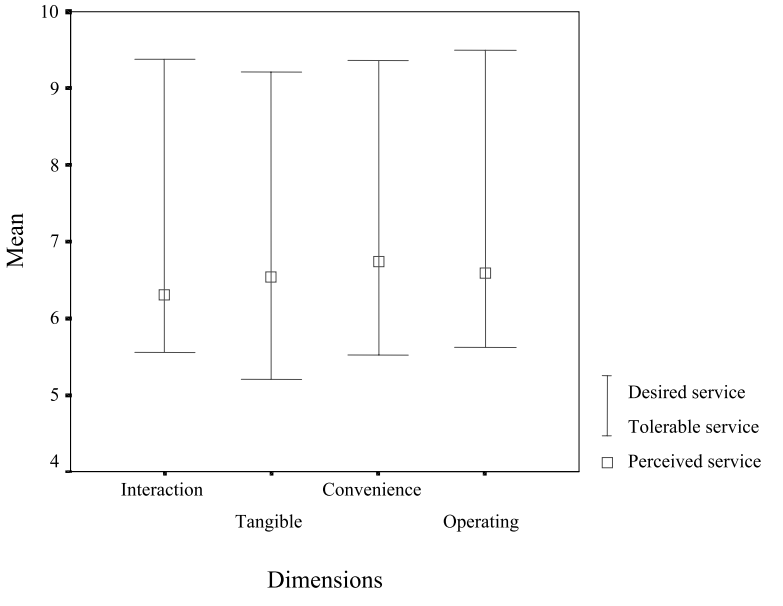

Figure 4. Tolerance range analysis

reduced to 50, which became 30 following further simplification. Although 23 questions remained after qualification, 20 items were selected following a final resimplification. The scale dramatically from the ten dimensions at stage 1, to only five. The dimensions at stage 3 were quite stable, with a combination of only two of the five. Regarding the items, although the original 64 were reduced to 20, the dimensions remained solid. Clearly restated, the scale that contained 20 items following a four-staged tiered development was as valid as that containing 30 questions. Cronbach's alpha indicates that the reliability of the final scale is the same with the scale formed following the first simplification (0.95). Our scale development procedure proved to be workable and demonstrates reasonable results. To develop their proprietary service quality scale, this procedure may be applied to other transportation industries, such as mass rapid transit, coach, freeway bus service or air service.

The items of the questionnaire were designed according to the conditions of the case. Some items mentioned in other studies or surveys were kicked off in this research during the developing stages due to the inapplicability to the present case. For example, the passengers in Taipei care mainly whether they can get on the bus, not whether they have a seat. The district of Taipei is not large in area, thus the distance of each bus route and the passenger's riding time are not very long. In addition, though there are few intercity coach stations, the Taipei city bus stops are much more in number and have a wide distribution (more than 2000 stops). The distance one has to walk from home to the bus stop is not a major issue and was discussed in focus group interviews. Besides, the bus speed in Taipei is restricted to 40 km/h by the TCG due to safety reasons. Therefore, the average speed of a bus cannot be determined by passengers. Finally, it is quite cheap for Taipei citizens to take a bus, because the low price policy of the TCG means that the fare for city buses has not risen over 8 years. The fare of each Taipei city bus is charged by section and each section costs NTD15 (i.e. about US$0.47 or £0.27). Ordinary passengers can mostly reach their destination by

taking one or at most two sections. In sum, although the service quality scale developed by this study has not included all indicators of bus service quality, it can fit the phenomenon of the city bus business in Taipei. In other words, this instrument is really made to measure according to one specific city.

According to the PZB's concept, customers' expectation of a bus service in this study is divided into two different levels: the desired service level and the tolerable service level. The distance between the desired service and the tolerable service is considered as the tolerance range. That is different from asking passengers directly about the importance of each service attribute. The implication of the tolerance range is that customers may accept variation within a range of performance (Liljander and Strandvik, 1993). The company must pay attention to the passengers and try to improve the service, while the passengers' perceived service is too close to the tolerable service but far from the desired service. This theme is already discussed materially in the field related to service quality research. Nevertheless, it is unusual to apply tolerance range analysis to the bus industry. The result of this study demonstrates that such a concept can really offer different and useful information and is worth carrying on in further research.

Despite several simplifications, the scale and dimensions remained highly reliable and valid. It is a good measurement tool for managers to assess a bus system's service quality. By assessing the scores of the four dimensions and items, managers can inspect whether or not their service quality is acceptable by passengers. If the result of assessment is unacceptable, managers may take corrective action such as changes in strategy, structure, compensation practices or training programmes, job redesign, or the replacement of personnel (Robbins and Coulter, 1999). Moreover, the service quality scale can also be employed over a longer period. According to the study results based on periodically tracking service quality trends, managers can monitor changes in perceptions and compare bus service performance over various periods. This could then be applied to develop new service strategies or improvement alternatives. However, what needs to be recognized is that while most attributes of service quality are controllable by the company's managers, some others are not. Driver behaviour, service equipment and the operating network, for example, are more directly the responsibility of the bus company than is unreliability, which is most often due to traffic congestion.

Some discussions and suggestions about sampling and related issues are as follows. The sample frames of two investigations in this paper are both similar to the previous relevant surveys in Taipei. However, the sample size of this research is restricted by the research time and budget. To improve the quality and representativeness of the materials, it is suggested that future research increase the sample size. Besides, to develop a reliable instrument needs not only a quantitative investigation, but also a qualitative interview to realize what the passenger really wants. To have better representation of the sample, researchers can apply the process of the multivariable sampling (Cunningham *et al.*, 2000). Thus, the methodological triangulation in this paper, which includes two questionnaires and one focus group interview, can supplement the difference of divergent samples. Moreover, the questionnaire data are collected about the bus stops or shelters in this study, since that is a feasible and practicable alternative in an exploratory research. It is suggested that a future study could conduct other ways of survey method, such as telephone interviews. Finally, this research only interviews regular bus users, whereas their requirements may be different to non-regular passengers' opinions due to a degree of familiarity with the bus service. It is also an important issue to

bus companies and needs further study to realize the reason why some citizens still choose other modes of transport rather than a city bus.

Acknowledgements

The authors acknowledge the support of the National Science Council of the Republic of China in Taiwan in funding the research. They are also thankful to the reviewers who provided very useful comments and guidance.

References

Alter, C. H. (1976) Evaluation of public transit services: the level of service concept, *Transportation Research Record,* 606, pp. 37–40.

Anderson, J. C. and Gerbing, D. W. (1988) Structural equation modeling in practice: a review and recommended two-step approach, *Psychological Bulletin,* 103(3), pp. 411–423.

AVV Transport Research Centre (2004) *Passenger Transport in the Netherlands* (Rotterdam: Ministry of Transport, Public Works and Water Management of Netherlands). Available at: http://www.rws-avv.nl/pls/portal30/docs/10197.PDF) (accessed 6 February 2006).

Babakus, E. and Boller, G. W. (1992) An empirical assessment of the SERVQUAL scale, *Journal of Business Research,* 24(3), pp. 253–268.

Bakker, J. J. (1976) Transit operating strategies and levels of service, *Transportation Research Record,* 746, pp. 41–47.

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

Bienstock, C. C., Mentzer, J. T. and Bird, M. M. (1997) Measuring physical distribution service quality, *Journal of the Academy of Marketing Science,* 25, pp. 31–44.

Bitner, M. J. (1990) Evaluating service encounters: the effect of physical surroundings and employee responses, *Journal of Marketing,* 54, pp. 69–82.

Botzow, H. (1974) Level of service concept for evaluating public transit, *Transportation Research Record,* 857, pp. 18–21.

Boulding, W., Kalra, A., Staelin, R. and Zeithaml, V. A. (1993) A dynamic process model of service quality: from expectation to behavioral intentions, *Journal of Marketing,* 30(1), pp. 7–27.

Bouman, M. and Van der Wiele, V. (1992) Measuring service quality in the car service industry: building and testing an instrument, *International Journal of Management,* 3(4), pp. 4–16.

Brown, T. J., Churchill, Jr, G. A. and Peter, J. P. (1993) Research note: improving the measurement of service quality, *Journal of Retailing,* 69, pp. 127–139.

Buchanan, C. (2005) *Bus Passenger Satisfaction Survey, 2004* (Edinburgh: Scottish Executive Social Research). Available at: http://www.scotland.gov.uk/Resource/Doc/57346/0016634.pdf) (accessed 6 February 2006).

Carman, J. M. (1990) Consumer perceptions of service quality: an assessment of the SERVQUAL dimensions, *Journal of Retailing,* 66(1), pp. 33–55.

Carter, D. N. and Lomax, T. J. (1992) Development and application of performance measures for rural public transportation operators, *Transportation Research Record,* 1338, pp. 28–36.

Churchill, G. A. (1979) A paradigm for developing better measure of marketing constructs, *Journal of Marketing Research,* 16, pp. 64–73.

Cronbach, L. J. and Meehl, P. E. (1955) Construct validity in psychological tests, *Psychological Bulletin,* 52, pp. 281–302.

Cronin, Jr, J. J. and Taylor, S. A. (1992) Measuring service quality: a reexamination and extension, *Journal of Marketing,* 56, pp. 56–68.

Cunningham, L. F., Young, C. E. and Lee, M. (2000) Methodological triangulation in measuring public transportation service quality, *Transportation Journal,* 40(1), pp. 35–47.

Dabholkar, P. A., Thorpe, D. I. and Rentz, J. O. (1996) A measure of service quality for retail stores: scale development and validation, *Academy of Marketing Science Journal,* 24(1), pp. 3–14.

Department for Transport of England (2005) *Bus Quality Indicators: England.* Available at: http://www.dft.gov.uk/stellent/groups/dft_transstats/documents/page/dft_transstats_041049.hcsp) (accessed 6 February 2006).

DeVellis, R. F. (1991) *Scale Development: Theory and Applications.* Applied Social Research Methods Series Vol. 26 (Newbury Park, CA: Sage).

Fielding, G. J. (1992) Transit performance evaluation in the USA, *Transportation Research,* 26A(6), pp. 483–491.
Fielding, G. J. and Anderson, S. C. (1983) Public transit performance evaluation: application to section 15Data, *Transportation Research Record,* 947, pp. 1–7.
Fielding, G. J., Babitsky, T. T. and Brenner, M. E. (1985) Performance evaluation for bus transit, *Transportation Research,* 19A(1), pp. 73–82.
Finn, D. W. and Lamb, Jr, C. W. (1991) An evaluation of the SERVQUAL scale in a retailing setting, *Advances in Consumer Research,* 18, pp. 312–322.
Fowler, Jr, F. J. (1995) *Improving Survey Question: Design and Evaluation.* Applied Social Research Methods Series Vol. 38 (Newbury Park, CA: Sage).
Hatcher, L. (1998) *A Step-by-Step Approach to Using the SAS System for Factor Analysis and Structural Equation Modeling,* 3rd printing (Cary, NC: SAS Institute Inc.)
Hensher, D. A. and Daniels, R. (1995) Productivity measurement in the urban bus sector, *Transport Policy,* 2(3), pp. 179–194.
Hensher, D. A., Stopher, P. and Bullock, S. (2003) Service quality—developing a service quality index in the provision of commercial bus contracts, *Transportation Research,* 37A, pp. 499–517.
Joreskog, K. G. and Sorbom, D. (1993) *LISREL8: User's Reference Guide* (Chicago, IL: Scientific Software International).
Kaiser, H. F. and Rice, J. (1974) Little jiffy mark IV, *Educational and Psychological Measurement,* 34, pp. 111–117.
Krueger, R. A. and Casey, M. A. (2000) *Focus Groups: A Practical Guide for Applied Research,* 3rd edn (Newbury Park, CA: Sage).
Liljander, V. and Strandvik, T. (1993) Estimating zones of tolerance in perceived service quality and perceived service value, *International Journal Service Industry Management,* 4(2), pp. 6–28.
McCormack, L. (1994) The qualitative to understanding service quality, *Asia Pacific Journal of Marketing and Logistics,* 6(12), pp. 63–80.
Mentzer, J. T., Flint, D. J. and Kent, J. L. (1999) Developing a logistics service quality scale, *Journal of Business Logistics,* 20(1), pp. 9–32.
Metropolitan Transit Authority of Harris County (1984) *Mass transit Evaluation Method* (Washington DC: UMTA).
Morgan, D. C. (1988) *Focus Groups as Qualitative Research* (Newbury Park, CA: Sage).
Parasuraman, A., Zeithaml, V. A. and Berry, L. L. (1985) A conceptual model of service quality and its implications for future research, *Journal of Marketing,* 49, pp. 41–50.
Parasuraman, A., Zeithaml, V. A. and Berry, L. L. (1988) SERVQUAL: a multiple-item scale for measuring consumer perceptions of service quality, *Journal of Retailing,* 64(1), pp. 12–40.
Parasuraman, A., Zeithaml, V. A. and Berry, L. L. (1991) Refinement and reassessment of the SERVQUAL scale, *Journal of Retailing,* 67(4), pp. 420–450.
Public Transport Council of Singapore (2005) *Findings of Bus Passenger Satisfaction Survey.* Available at: http://www.ptc.gov.sg/statistics_summary.asp) (accessed 6 February 2006).
Pullen, W. T. (1993) Definition and measurement of quality of service for local public transport management, *Transport Reviews,* 13(3), pp. 247–264.
Robbins, S. P. and Coulter, M. (1999) *Management,* 6th edn, ch. 17 (Englewood Cliffs, NJ: Prentice-Hall).
Shemwell, D. J. and Yavas, U. (1999) Measuring service quality in hospitals: scale development and managerial applications, *Journal of Marketing Theory and Practice,* 7(3), pp. 65–75.
Smith, A. M. (1999) Some problems when adopting Churchill's paradigm for the development of service quality measurement scales, *Journal of Business Research,* 46, pp. 109–120.
Stewart, D. W. (1990) *Focus Groups: Theory and Practice* (Newbury Park, CA: Sage).
TIS.PT, Consultores em Transportes, Inovação e Sistemas, S.A. (2003) *Handbook of Managing and Assessing Regulatory Evolution in Local Public Transport Operations in Europe.* Available at: http://www.tis.pt/proj/maretope/handbook.pdf) (accessed 6 February 2006).
Transport for London (2006) *Network Performance of London Buses.* Available at: http://www.tfl.gov.uk/buses/about/performance/service-report.asp) (accessed 6 February 2006).
Triplett, J. L., Yau, O. H. M. and Neal, C. (1994) Assessing the reliability and validity of SERVQUAL in a longitudinal study: the experience of an Australian organization, *Asia Pacific Journal of Marketing and Logistics,* 6(12), pp. 41–62.
Wipper, L. R. (1993) Performance measurement: producing results at the Oregon department of transportation, *Transportation Research Record,* 1395, pp. 168–175.
Zeithaml, V. A., Parasuraman, A. and Berry, L. L. (1990) *Delivering Quality Service: Balancing Customer Perceptions and Expectations* (New York: Free Press).