CIRRELT

Centre interuniversitaire de recherche
sur les réseaux d'entreprise, la logistique et le transport

Interuniversity Research Centre
on Enterprise Networks, Logistics and Transportation

Measuring Quality of Service in Dial-a-Ride Operations : The Case of a Canadian City

Julie Paquette
François Bellavance
Jean-François Cordeau
Gilbert Laporte

March 2011

CIRRELT-2011-19

Bureaux de Montréal :
Université de Montréal
C.P. 6128, succ. Centre-ville
Montréal (Québec)
Canada H3C 3J7
Téléphone : 514 343-7575
Télécopie : 514 343-7121

Bureaux de Québec :
Université Laval
2325, de la Terrasse, bureau 2642
Québec (Québec)
Canada G1V 0A6
Téléphone : 418 656-2073
Télécopie : 418 656-2624

www.cirrelt.ca

Measuring Quality of Service in Dial-a-Ride Operations : The Case of a Canadian City

Julie Paquette[1,2,*], François Bellavance[1,3], Jean-François Cordeau[1,2], Gilbert Laporte[1,3]

[1] Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation (CIRRELT)

[2] Department of Logistics and Operations Management, HEC Montréal, 3000 Côte-Sainte-Catherine, Montréal, Canada H3T 2A7

[3] Department of Management Sciences, HEC Montréal, 3000 Côte-Sainte-Catherine, Montréal, Canada H3T 2A7

Abstract. In many countries, dial-a-ride services are provided by public authorities to elderly and handicapped people who cannot use regular transit. Cost minimization is key to running these services, but one can observe a growing interest in quality improvement. A first step in improving quality is to define a quality measurement scale. A second step is to incorporate quality measurements in mathematical models that serve as a basis for optimization algorithms. To develop such a measurement instrument, an extensive survey of dial-a-ride users was conducted in Longueuil, the largest suburb of Montreal, Canada. This paper describes the steps of the survey and presents the results and its main conclusions.

Keywords. Dial-a-ride services, quality of service, measurement scale.

Acknowledgements. This work was supported by the Social Sciences and Humanities Research Council of Canada (SSHRC) and by the Natural Sciences and Engineering Council of Canada (NSERC) under grants 227837-09 and 39682-10. This support is gratefully acknowledged

---

* Corresponding author: Julie.Paquette@cirrelt.ca

Dépôt légal – Bibliothèque et Archives nationales du Québec
Bibliothèque et Archives Canada, 2011

1 Introduction

Dial-a-ride services are typically provided to elderly and handicapped people who cannot use regular transit. These adapted services are usually performed by minibuses and taxis, and the routing and scheduling of the vehicles can be optimized by applying operations research algorithms. The typical chain of operations in adapted transportation services is depicted in Figure 1. As shown, users of the service must call 24 hours in advance to reserve their ride. An appointment is then confirmed for the next day. On the day of the appoinment, an adapted vehicle will fulfill the transportation request at the time requested. Since the service is public and often shared, there may be many users on board the vehicle at the same time.

The aging of the population will have a direct impact on dial-a-ride services, and will increase their demand, especially if little is done to adapt regular public transit services. Moreover, communities are becoming increasingly sensitized to equity and quality of living issues, particularly when it comes to the elderly and the handicapped. Dial-a-ride services are essential to the mobility of this segment of the population and are often publicly funded. Traditionally, the managers of these services have been concerned with costs issues as opposed to quality of service because they have to operate within tight budgets. However, one can sense an increased interest toward quality of service issues. This trend should benefit the users who often have no other transportation alternatives.

Quality is a multidimensional construct for which several definitions exist. Grönross (1984) was among the first to develop a specific definition of quality for the service sector. He defines quality as "the outcome of an evaluation process, where the consumer compares his expectations with the service he perceives he has received" (p.37). A drawback of this definition, called the disconfirmation paradigm, is that expectation is a difficult concept to operationalize because it can be interpreted in many ways. As proposed by Schneider and White (2004), we will synthesize the ways to define quality under three approaches: philosophical, technical and customer-based quality. Philosophical quality is synonym of excellence and can only be recognized when seen. Technical quality refers to the conformity to specifications used by the provider of the service to set a level of quality. Usually, it can be easily measured because it is based on objective criteria. Finaly, customer-based quality is based on the perceptions of users for different dimensions of quality and is thus considered to be more subjective.

Several models have been proposed to define quality and most work with different dimensions. Grönross (1984) distinguishes between technical quality which refers to the result of the service, and functional quality which refers to the process or service experience. Rust and Oliver (1994) have developed a model that incorporates three dimensions: customer-employee interaction, outcomes and service environment. Finally, the best known quality model is the one developed by Parasuraman et al. (1985) which identifies five gaps and five dimensions: tangibles, responsiveness, reliability, assurance and empathy. Seth et al. (2004) have reviewed and commented on nineteen existing models, but their survey ignores the hierarchical model developed by Brady and Cronin (2001) in which the three models just described are merged.

To measure quality of service, some researchers have proposed to use statistical methods commonly used in the manufacturing sector, but these only apply to some tangible and easily measurable service components. In response to this, Parasuraman et al. (1988) have developped the SERVQUAL measurement scale. The SERVPERF version of the scale only uses the perceptions and does not use the disconfirmation paradigm as is done in the SERVQUAL. Several researchers have criticized SERVQUAL on theoretical, methodological and psychometric grounds. According to Powpaka (1996) and others, this scale is not efficient for some services, which explains why many researchers prefer constructing measurement scales not based on SERVQUAL or adapting it to specific services.

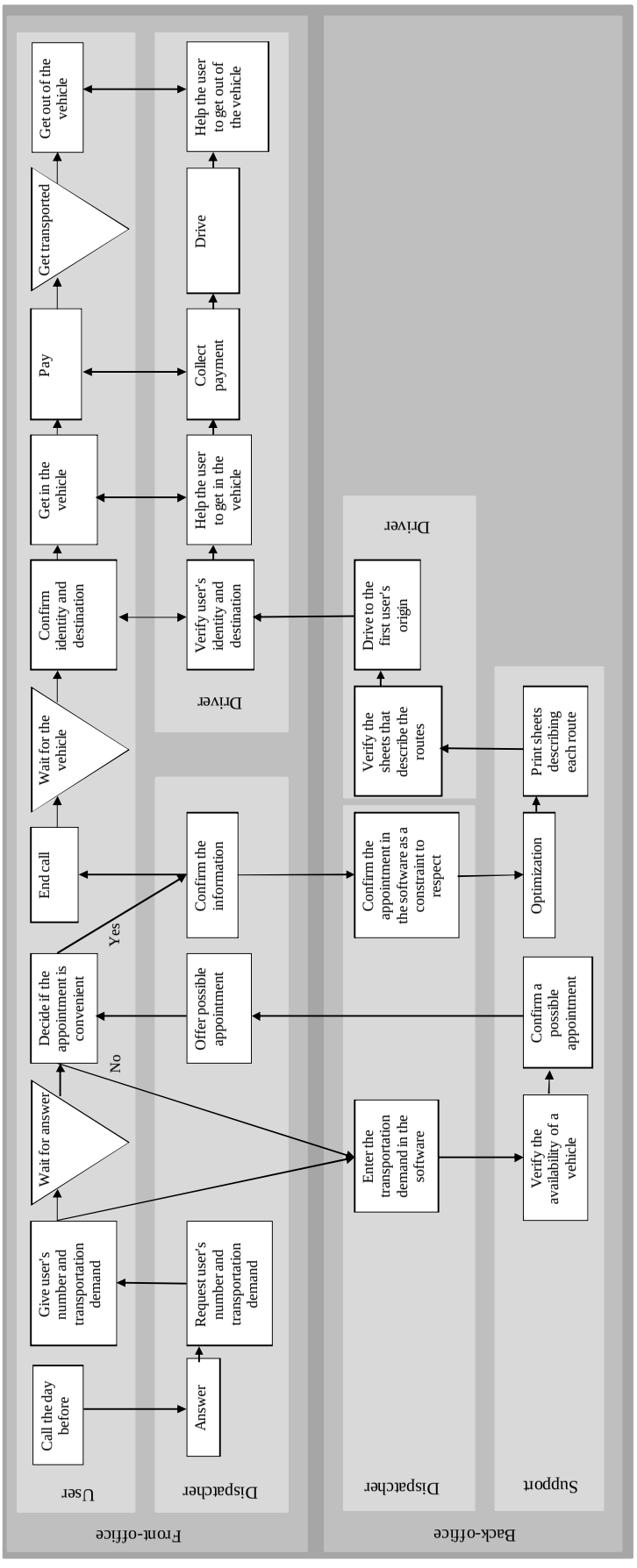

Figure 1: Chain of operations in adapted transportation services

Quality of service in dial-a-ride operations has not been extensively studied. Only a few articles exist on the measurement of quality in these services and none offers workable tools to improve it. A complete literature review is presented in Paquette et al. (2009). The three main studies described in this survey are summarized here. Pagano and McKnight (1983) were the first to develop a quality measurement scale for dial-a-ride services based on a list of attributes originally established for public transportation. A questionaire including these attributes was sent to 659 dial-a-ride users and the authors concluded that providing a high quality dial-a-ride service is difficult, since quality in this context depends on several attributes and quality criteria are user-dependent. A drawback of their study is that users were not asked to define their needs and expectations in a qualitative research phase prior to the questionaire. Thus, the dimensions reported could be based on preconceptions and could be biased. Moreover, this study excludes the blind, the deaf and the mentally handicapped from the sample, which may have created a bias in terms of representativity. A second study was performed by Denson (2000) who combined the quality attributes of Pagano and McKnight (1983) in a questionaire administered by phone to 2500 users of dial-a-ride services and identified factors of user satisfaction and dissatisfaction. Finally, a third study by Knutsson (1999) in which quality attributes were used to estimate the demand for public dial-a-ride services was conducted in Sweden. A response rate of 65% allowed Knutsson to highlight six quality attributes that were the most important to the users. These formed the basis of a utility function which was incorporated in a Logit model used to predict the users' willingness to pay for the service.

In another vein, the operations research literature contains several articles on the dial-a-ride problem (see, e.g., the survey by Cordeau and Laporte (2007)). The solution to this problem is a set of vehicle routes and schedules that minimize operational costs and user inconvenience, typically handled through the imposition of time windows or ride time constraints. However, beyond the imposition of these two classes of constraints, quality of service criteria are ignored in most operations research studies.

To summarize, improving quality of dial-a-ride services is likely to gain in importance in the near future. However, this aspect has only received limited attention in the quality of service literature and in operations research. As a rule, minimizing operating costs is the sole objective of most algorithms, and quality of service is only handled indirectly through the imposition of constraints. We believe that to properly integrate quality of service within optimization algorithms, a valid and reliable quality measurement scale should first be defined and then incorporated within a multi-criteria optimization process.

The purpose of this study is to provide analytical tools to (1) measure quality of service in dial-a-ride services based on users perceptions, (2) identify which attributes are the most important, and (3) identify control variables by which subgroups of users having different perception levels can be differentiated. To this end, we have conducted an extensive survey of dial-a-ride users in the conurbation of Longueuil, the largest suburb of Montreal, Canada. This paper describes the steps of the survey and its main results and conclusions.

The remainder of this paper is organized as follows. Section 2 describes the characteristics of the population under study and the methodology used to design the questionaire and the sampling scheme. Section 3 presents the results of an exploratory factor analysis performed to identify the main dimensions of quality in dial-a-ride services. Conclusions are provided in Section 4.

2 Methodology

We will now describe the methodology of our study.

2.1 Description of the population under study

This study was conducted in cooperation with the Réseau de Transport de Longueuil (RTL) which operates dial-a-ride services in Longueuil and adjacent areas. The conurbation of Longueuil, which has 397 000 inhabitants, is located on the south shore of Montreal. The RTL provides dial-a-ride services to a pool of 2600 users. On an average day, 450 to 550 people use these services. We were granted permission by a government agency to conduct a survey with the users. We were provided with a full list of active users, including their name, sex, address, phone number and information on whether they are mentally handicapped or not. An active user is defined as one who has used the service at least once during the last year at the time the survey was conducted (February 2008). This list excludes minors and in the case of mentally handicapped users, we dealt with their tutor who is deemed to be an appropriate respondent because this is the person who calls the reservation center and is aware of the perceptions of the user.

Some of the potential respondents were excluded because they had died, had moved or did not use the service anymore. In contrast to the studies of Pagano and McKnight (1983) and of Denson (2000), the mentally handicapped and the blind were not excluded from our study. This is rather important because these users represent a significant proportion of the population (about 30%).

2.2 Data collection methods

According to Lovelock et al. (2009), researchers agree that the distinctive nature of services requires a different approach in defining and measuring quality of service. We have therefore concentrated on identifying the perceptions and needs of users in order to determine the attributes used to explain and measure the quality of dial-a-ride services. A combination of two methodologies was selected, based on the recommendation of researchers who had criticized the SERVQUAL, or of researchers who had performed quality studies on other types of services (Martinez Caro and Garcia (2007), Karatepe et al. (2005), Schneider and White (2004), and Devlin and Dong (1994)). Our intention was to first consult a focus group of users to help determine a list of relevant attributes, before proceeding with a postal questionaire survey. However, ethical considerations prevented us from using focus groups. Indeed, a user focus group was judged to be too risky by our ethics committee, in the sense that participants could be identified. We therefore resorted to conducting individual interviews with a sample of users in an initial phase. Interviews with the manager of the dial-a-ride service and with the director of the users' association were also performed to verify that no important attributes were missing.

Based on the results of the initial phase, a questionaire was designed and sent by mail to a large sample of users, except the blind and people with motor disability who could answer it by phone. Because the population under study is well identified and users feel concerned about the subject of the study, it was concluded that a postal questionaire could be used even if this collection method can sometimes generate a low response rate. An Internet questionaire could not be considered because about half of the population under study is older than 65 and probably does not use this technology (a posteriori, 65% of the survey participants have indeed declared having no access to the Internet), and it would have been too costly and time consuming to fill in the questionaire with each respondent over the phone. Using a postal questionaire allowed us to ask more questions than would have been possible by phone. This was important to us because we wanted to collect as much information as possible.

2.3 Questionaire design

Because the purpose of this study is threefold, the postal questionaire has three parts, each with a specific goal. The first part of the questionaire aims at measuring quality of service perceived

by users. The first step is to identify the attributes of quality of dial-a-ride services. Thus, in an initial phase semi-structured interviews with dial-a-ride users were used to determine the attributes on which to base the construction of the quality measurement scale specific to dial-a-ride services and to complement preconceptions held by the researchers and the service manager. Sampling for this part of the study was done from the list of active users. Three attempts to reach each randomly selected user were made. If the user could not be reached, a new one was randomly chosen from the list. Each interview was conducted at the user's home to ensure that the participant would not have to use the dial-a-ride service and thus prevent possible identification of the user by the provider. The resulting sample is quite heterogeneous in terms of sex, area of residence, type of handicap, and age; it thus follows the criteria set by Arnould and Epp (2006). The number of interviews performed was based on the saturation level criterion. After the 15th interview, the process was stopped because no new information was generated.

The interviews were taped and transcribed to enable a better analysis. We have followed the procedure suggested by Richards (2005) to analyze the data. Themes were first extracted from the transcripts which were then read repeatedly, and a list of attributes was constructed from these themes, with approximately the same level of detail for each theme. This list was then compared to the set of attributes suggested by Pagano and McKnight (1983), Knutsson (1999), the SERVQUAL and a previous study performed by the service provider. Some attributes were similar to those used by other researchers, but a few were specific to our study. Also, some of the attributes used in previous studies did not apply to our case.

The list was then updated to make the categories consistent and to include some attributes that did not emerge in the interviews but were considered in the literature. The resulting list was submitted to the operations manager and to the director of the users' association, which resulted in further marginal adjustments. The final list contains 56 attributes of the quality of service specific to dial-a-ride services. These attributes were then used to construct the postal questionaire, in which a question is asked for each of them. The data collected with the first part of the questionaire should enable us to identify the dimensions of quality of service in dial-a-ride operations. The data should also help the provider identify its strengths and weaknesses. All questions are declarative sentences. We have chosen to measure perceptions and not use the disconfirmation paradigm partly for the sake of brevity, and also because Carrillat et al. (2007) have concluded that the SERVPERF is equivalent to the SERVQUAL in terms of validity and relevance. A question on the global quality perceived was also added to this part of the questionaire.

As in the SERVQUAL, we have used a Likert scale. Also as in Parasuraman et al. (1988), we have made the hypothesis that there exists a continuous variable underlying the scale, which allows for analysis as if it were a quantitative scale. We have chosen a ten-level scale (from 1: completely disagree to 10: completely agree) which does not offer a central choice and therefore forces the respondent to choose between the negative and the positive side (Bishop, 1987).

The second part contains questions about the importance of each attribute for the users. Here, we did not use regression analysis as was done in the SERVQUAL (indirect method). Indeed, a direct ranking method was preferred because as reported by Oh (2001), "Neslin (1981) showed that statistically derived relative importance had superior predictive validity than self-reported absolute importance." To facilitate the task of the respondent, the attributes were subdivided into five groups of five attributes each, as suggested by Abalo et al. (2007). It was thus easier to identify the most important attributes within each group, and then among the five attributes chosen across all groups. The data collected with this question should help identify the attributes that are currently considered to be the most important. It is expected that improving the quality of a service attribute considered to be important should lead to a decrease in the relative importance of this attribute in the future. Thus, this information is dependent on the performance of the provider, as Sampson and Showalter (1999) have found. We have also added to this part of the questionaire a question on the relative

importance of measurable attributes. This information should help rank the attributes that can be easily incorporated when designing the routing and scheduling of the vehicles.

Finally, the third and last part of the questionaire contains socio-demographic questions which help determine whether different groups of users have different perceptions, expectations or needs. Two questions were also added asking the users whether they had made a complaint to the adapted transportation service or had suffered at least one delay of more than 30 minutes in the last month. These two questions should allow us to verify the validity of the measurement instrument.

The questionaire was pre-tested with the director of the users' association and was then sent to eight users previously reached by telephone and who had agreed to answer it. Six of them effectively filled and returned the questionaire and accepted to later criticize it by phone. Minor changes were made and some details were added to the instructions.

2.3.1 Sampling and response rate

To reach as many users as possible, while keeping the costs and timeframe within reasonable limits, the following approach was privileged. First, a random sample of 1175 active users were contacted by phone and asked whether they would accept to participate in the study. If so, the postal questionaire was mailed to them or was filled over the phone. If not, their name was dropped from the list. Each user in the sample was contacted up to five times on different days and times of day to maximize the contact rate. To improve the response rate, a reminder was sent three weeks after the first mailing of the questionaire.

The sampling unit is a user, and as mentioned before, the sample was extracted from the active users list. A stratified sampling method was used to ensure the statistical efficiency of the sample. Thus, the respondents were randomly selected from strata formed on the basis of their borough of residence and of their sex. It would have been useful to stratify the population of users by age or type of handicap, but these data were not available in the database. The stratified sampling method used is proportional, meaning that the number of users sampled in each strata is proportional to the number of users having the characteristic in the population.

From the initial sample of 1175 users (45% of the total population size), 235 were not part of the population under study because they had moved, had died or did not use the service anymore. From the remaining 940 active users in the sample, only 83 could not be reached, yielding a contact rate of 92% (857/940). Among the users who were reached, 572 accepted to respond to the mailed questionaire or by phone, which translates into an acceptance rate of 66.7% (572/857). Of the 572 questionaires sent, 333 were returned. We note that the remainder had a positive impact on the return rate. In addition, among the 333 questionaires received, two had to be discarded because they were unusable. The overall response rate is therefore 38.6% (331/857), which we view as very satisfactory.

To ensure that respondents are representative of the population, we have compared the proportion of various user groups in the population with their proportion among the respondents, as presented in Table 1. The data on the population is based on aggregated data provided in an internal report sent once a year by the service provider to the government. The table brings out some differences between subgroups of users in the population and subgroups of users who responded to the questionaire, but overall we can conclude that respondents adequately represent the population under study.

Table 1: Comparison between the users in the population and in the sample

Population ($N = 2561$)			
	Age		
Type of handicap	[21,65]	[66,+]	Total
Movement disorder or organic disability, wheelchair	15%	16%	31%
Movement disorder or organic disability, ambulatory	11%	24%	35%
Intellectual disability	17%	1%	18%
Psychical disability	1%	0%	1%
Visual impairment	3%	3%	6%
Others	3%	6%	9%
Total	50%	50%	100%
Respondents ($n = 331$)			
	Age		
Type of handicap	[18,64]	[65,+]	Total
Movement disorder or organic disability, wheelchair	14%	16%	30%
Movement disorder or organic disability, ambulatory	12%	18%	30%
Intellectual disability	14%	1%	15%
Psychical disability	0%	1%	1%
Visual impairment	5%	3%	8%
Others	9%	7%	16%
Total	54%	46%	100%

Some attributes present a low rate of response because they were not applicable to all users. This is sometimes the case of attributes related to technology, vehicles and customer service. When a respondent answers that these features do not apply, they are treated as missing data.

3 Results of the questionaire survey

This section describes the results of our mailing questionaire survey and the various statistical analyses we have performed using SAS 9.1 for Windows.

3.1 Quality dimensions in dial-a-ride services

An exploratory factor analysis was used to extract the dimensions of the quality construct measured with the 56 attributes. Each attribute was assigned to a dimension and the unnecessary ones were removed (i.e. attributes with a factor loading less than 0.3 on all dimensions were removed). The extraction of factors was done using the iterated principal factor analysis (method PRINIT in SAS) with a varimax rotation. As mentioned, the relatively high rate of "not applicable" responses for certain blocks of attributes (i.e. technology, vehicles and customer service) has prevented us from using these attributes in the main factor analysis; these data were analyzed separately.

A preliminary analysis has enabled us to conclude that there exist correlations between many of the attributes. Moreover, the measure of sampling adequacy (MSA), which is equal to 0.9082 for our sample, shows it is adequate for a factor analysis. Two criteria were used to determine the number of factors: the eigenvalues and the scree plot. There are seven eigenvalues greater than one, and the scree plot stabilizes after four or eight factors. We have therefore performed our factor analysis with four, five, six, seven and eight factors. Table 2 presents the root-mean-square residuals (RMSR) for each of these common factor models. The easiest model to interpret has proved to be the model with

eight factors (note that the eighth eigenvalue is equal to 0.94 which is close to 1). Table 3 presents the loadings of each attribute when eight factors are used. A loading with an asterisk means that it is greater than 0.3. When an attribute has a loading greater than 0.3 for more than one factor, it is associated to the factor for which the loading is the largest (loading in bold in Table 3). The eight (rotated) factors explain 66.5% of the total common variance. Based on this model, scales were constructed by taking the average score of the attributes for each factor. Table 4 summarizes the eight dimensions of quality and their related attributes. For each dimension, the proportion of variance explained and the Cronbach's alpha coefficient are presented in parentheses. Moreover, because the Cronbach's alpha of each scale is well above 0.6, we can conclude that the internal consistencies of the scales are very satisfactory.

Table 2: Root mean square residuals (RMSR) for different number of factors retained in the model

Number of factors in the model	RMSR value
4	0.0511
5	0.0445
6	0.0379
7	0.0322
8	0.0280

Three other factor analyses were performed on the blocks of variables related to customer service ($n = 213$), technologies ($n = 112$) and vehicles ($n = 100$). The resulting dimensions of the quality of service and their related attributes are summarized in Table 5. For each dimension, the proportion of the variance explained and the Cronbach's alpha are mentionned in parentheses. First, the factor analysis conducted on the four attributes related to customer service has helped identify a single factor with an RMSR of 0.0142. The Cronbach's alpha coefficient associated to the scale is 0.9345, which leads to the conclusion that the internal consistency is very satisfactory. Moreover, when a_{38} is removed, the coefficient stays basically the same: 0.9333. That attribute was therefore not retained.

Another common factor analysis was performed on the block of attributes related to technology. First, the correlation matrix contains entries larger than 0.3. However, MSA = 0.5961 which is deemed to be "very poor". This may be explained by the fact that only three variables are included in this analysis. Performing the factor analysis results in a model with a single factor, yielding a Cronbach's alpha of 0.8580. Moreover, when the attribute related to the Internet web site is removed, the coefficient climbs to 0.9033.

Regarding the block of fifteen attributes on the vehicles, it is possible to conduct a factor analysis because there exist significant correlations between variables, and the MSA value is equal to 0.8093. Considering the eigenvalues criterion, the three-factor model was chosen and scales were constructed accordingly. The Cronbach's alpha coefficient is 0.9587 for the first scale (taxis), 0.8805 for the second (minibus), and 0.9071 for the third (taxi cleanliness) if the two attributes related to taxis seatbelts are removed.

We conclude from this analysis that there are 13 factors, and therefore the quality of adapted transportation services can be measured with as many different dimensions. Our study suggests the existence of more factors than that of Parasuraman et al. (1988). The dimensions "tangibles", "responsiveness", "reliability", "assurance" and "empathy" can all be associated to one or more of the dimensions found in our study. Some attributes are significant in more than one factor, which leads us to think that some of the dimensions found by the factorial analysis could be grouped under broader dimensions. For example, the "interaction" dimension could include the three dimensions corresponding to the three types of employees in contact with the users: drivers, dispatchers and customer service, and the dimension "information" could include the three separate dimensions related to it: "information is accurate", "information availability" and "speed to obtain information".

Table 3: Loadings ($\times 100$) associated to the attributes for the common factor model with eight factors ($n = 174$)

Attribute	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8
	Dispatchers	Drivers	Service configuration I	Experience of service	Information is accurate	Service configuration II	Information availability	Speed to obtain information
a_1	**76** *	6	13	4	23	0	18	11
a_2	**64** *	13	26	5	33 *	−2	10	15
a_3	**64** *	26	29	17	18	0	2	16
a_4	**77** *	15	15	16	3	11	30 *	12
a_5	**76** *	15	22	17	−1	12	23	18
a_6	**59** *	11	4	22	7	28	0	10
a_7	**80** *	31 *	14	20	3	24	6	−5
a_8	**80** *	19	20	24	10	15	10	7
a_9	10	**43** *	20	24	0	18	12	37*
a_{10}	21	**78** *	10	17	−1	2	16	10
a_{11}	14	**79** *	15	15	19	6	15	−9
a_{12}	14	**76** *	11	7	17	12	6	−5
a_{13}	3	**64** *	2	4	11	7	−3	8
a_{14}	14	**53** *	19	21	20	6	24	15
a_{15}	19	**48** *	19	9	8	21	−8	20
a_{16}	29	**74** *	20	26	−1	−1	4	13
a_{17}	12	10	**53** *	28	3	12	1	3
a_{18}	21	13	**69** *	9	6	10	16	19
a_{19}	20	16	**88** *	13	9	10	13	6
a_{20}	19	20	**52** *	13	7	23	7	21
a_{21}	29	29	**51** *	45 *	10	6	16	8
a_{22}	38 *	28	24	**54** *	28	18	11	7
a_{23}	34 *	27	30 *	**56** *	30	16	8	20
a_{24}	31 *	27	25	**64** *	17	10	7	18
a_{25}	29	40 *	41 *	**51** *	13	13	14	10
a_{26}	24	32 *	32 *	**59** *	15	12	20	10
a_{27}	14	26	6	19	**74** *	10	20	13
a_{28}	30 *	22	13	18	**69** *	10	2	7
a_{29}	19	15	18	7	5	**77** *	8	21
a_{30}	18	12	20	15	12	**69** *	14	0
a_{31}	28	14	12	11	11	15	**60** *	3
a_{32}	42 *	19	31 *	16	14	11	**71** *	3
a_{33}	23	13	35 *	13	10	12	12	**66***
a_{34}	38 *	12	8	17	27	14	−12	**53***
% variance explained	17%	13.7%	9.9%	7.6%	5.2%	4.8%	4.3%	4%

Table 4: List of the eight dimensions of the quality of service in dial-a-ride operations and their related attributes resulting from the first common factor analysis ($n = 174$)

Dispatcher (% variance explained = 17%; Cronbach's alpha = 0.9462)	Mean	SD
a_1 Answers my questions rapidly.	8.11	2.40
a_2 Is familiar with the provider's standards.	8.52	2.11
a_3 Understands the special needs of the users.	8.03	2.50
a_4 Is courteous and kind.	8.67	2.05
a_5 Shows by his attitude that he wants to help the user.	8.38	2.25
a_6 Accommodates the user as much as possible by allowing last minute changes.	7.23	3.16
a_7 Tries to meet the needs of the users.	8.03	2.60
a_8 In general, I would say I have an excellent relationship with the dispatchers.	8.44	2.22
Driver (% variance explained = 13.7%; Cronbach's alpha = 0.9216)		
a_9 Is courteous and kind.	8.62	2.00
a_{10} Knows the territory served by the provider.	8.63	1.87
a_{11} Shows by his attitude that he understands the special needs of the users.	8.38	2.16
a_{12} Drives safely.	8.24	2.41
a_{13} Makes sure that the seat belts are properly fastened.	8.38	2.59
a_{14} Helps the user get on and off the adapted transportation vehicle.	8.63	1.99
a_{15} Shows by his attitude that he wants to help the users.	8.65	2.33
a_{16} In general, I would say I have an excellent relationship with the drivers.	8.76	1.96
Service configuration I (% variance explained = 9.9%; Cronbach's alpha = 0.8670)		
a_{17} The dial-a-ride service provider serves a territory that meets the user's travel needs.	8.77	2.23
a_{18} The service hours for the dispatch service meet the needs of the users.	8.77	2.08
a_{19} The service hours of the adapted transportation service meet my needs.	8.96	1.88
a_{20} The deadline to book an occasional ride is reasonable.	8.57	2.26
a_{21} The dial-a-ride service meets the user's needs and enables them to be more mobile.	8.90	1.86
Experience of service (% variance explained = 7.6%; Cronbach's alpha = 0.9105)		
a_{22} The dial-a-ride service provider tries to keep the user's waiting time to a minimum.	7.87	2.41
a_{23} The dial-a-ride service provider respects the promised time slot.	7.91	2.37
a_{24} The dial-a-ride service provider tries to keep the user's ride time as short as possible.	8.06	2.37
a_{25} The dial-a-ride service provider tries to provide users with a pleasant ride.	8.67	1.99
a_{26} Users always have a very good experience when they use the dial-a-ride service.	8.23	2.32

Continued on next page

Information is accurate (% variance explained = 5.2%; Cronbach's alpha = 0.8588)	Mean	SD
a_{27} The provider's customer service department gives users accurate information.	7.63	2.55
a_{28} The dispatcher gives me accurate information.	7.50	2.63
Service configuration II (% variance explained = 4.8%; Cronbach's alpha = 0.8359)		
a_{29} An occasional ride must be booked no more than three days ahead of time, which is reasonable.	7.63	3.04
a_{30} The deadline to book a ride outside the territory is reasonable.	7.52	3.10
Information availability (% variance explained = 4.3%; Cronbach's alpha = 0.8564)		
a_{31} The dial-a-ride service provider provides the users with all the information they need on the service's operating rules described in the user's guide.	8.66	2.38
a_{32} The dial-a-ride service provider gives the users all the information they need on the service's operating rules when they request it by phone.	8.94	2.00
Speed to obtain information (% variance explained = 4%; Cronbach's alpha = 0.7689)		
a_{33} The user can reach the provider's customer service department easily and quickly.	8.24	2.36
a_{34} The user can reach the dispatcher easily and quickly.	7.82	2.58

The "tangible" dimension could also incorporate three dimensions (minibuses, taxis and cleanliness of vehicles), and the dimension "service configuration" could agregate the two dimensions related to it. Thus, a hierarchical model such as the one developed by Brady and Cronin (2001) could probably be appropriate in our context. However, this analysis cannot be performed because of missing data (see Section 2.3.1).

3.2 Reliability and validity verification

The reliability of the questionaire was tested in the previous section. Indeed, the Cronbach's alpha coefficients, calculated for each of the dimensions (factors) listed, were always very high, which leads to the conclusion that the measure instrument is reliable. Recognizing that reliability is a necessary but not sufficient condition for validity, it is necessary to perform further analyses.

To assess content validity, the answers to the section "other comments" of the questionaire were reviewed to ensure that no aspect addressed in that section was missing from the questionaire. To assess content validity, it is also possible to rely on the existing theory about quality. Thus, since the attributes were drawn from interviews with users, were reviewed in comparison with the literature, and were verified by the users' association manager and the service manager, it was concluded that no significant aspect of the service had been excluded from the study.

On the other hand, it is also necessary to test the criterion-related validity. To this end, we have included in the questionaire a single question about the overall quality of service. The correlations between the results to this question and the score of each of the 13 factors are presented in Table 6. The results of this exercice are conclusive. Indeed, the p-values for all the correlations are less than 0.0001 and all the correlations are greater than 0.45. These results thus show a very good level of criterion-related validity.

Table 5: List of the five dimensions of the quality of service in dial-a-ride operations and their related attributes resulting from the three separate common factor analyses for customer service ($n = 213$), technologies ($n = 112$), and vehicles ($n = 100$)

Customer service (% variance explained = 82.4%; Cronbach's alpha = 0.9333)	Mean	SD	Loading (×100)
a_{35} The dial-a-ride service provider shows a sincere interest in getting the user's problem resolved.	7.68	2.84	91
a_{36} The dial-a-ride service provider follows up on my complaints.	7.68	2.83	92
a_{37} The employee at customer service can resolve the user's problem quickly.	7.39	2.91	89
Technology (% variance explained = 82.5%; Cronbach's alpha = 0.9033)			
a_{39} The software the dispatchers use to reserve my ride enables them to meet my needs.	7.91	2.67	91
a_{40} The interactive voice response system (IVR) meets the user's needs.	7.65	2.84	91
Taxis (% variance explained = 38.2%; Cronbach's alpha = 0.9587)			
a_{42} The taxis are comfortable.	8.72	2.02	83
a_{43} The interior design of the taxis meets my needs.	8.48	2.26	78
a_{44} The interior design of the taxis makes it easy to get in and out.	8.20	2.40	91
a_{45} The accessible taxis are comfortable.	8.44	2.29	86
a_{46} The interior design of the accessible taxis meets my needs.	8.34	2.48	88
a_{47} The interior design of the accessible taxis makes it easy to get in and out.	8.36	2.38	91
Minibus (% variance explained = 25.8%; Cronbach's alpha = 0.8805)			
a_{48} Are comfortable.	7.37	2.81	66
a_{49} Are clean.	8.54	2.04	70
a_{50} The interior design of the minibus meets my needs.	8.52	2.27	91
a_{51} The interior design of the minibus makes it easy to get in and out.	8.66	2.17	91
a_{52} The anchors and seatbelts work properly.	9.03	1.87	81
Taxis cleanliness (% variance explained = 14.6%; Cronbach's alpha = 0.9071)			
a_{53} The taxis are clean.	8.68	1.95	84
a_{54} The accessible taxis are clean.	8.53	2.19	73
Variables that were not kept in any dimensions			
a_{38} In general, I would say that I have an excellent relationship with the employees at the customer service.	8.53	2.16	n.a.
a_{41} The service provider's Internet site meets my needs.	7.01	3.36	n.a.
a_{55} The seatbelts in the taxis work properly.	9.04	1.84	n.a.
a_{56} The anchors and seatbelts in the accessible taxis work properly.	8.87	2.00	n.a.

Table 6: Correlations between the perception of the overall quality of dial-a-ride services and the 13 factors

Factor	Correlation with perceptions of global quality
Dispatchers	0.7195
Drivers	0.6432
Service configuration I	0.6464
Experience of service	0.7713
Information is accurate	0.5883
Service configuratio nII	0.4513
Information availability	0.4768
Speed to obtain information	0.5251
Customer service	0.5914
Technologies	0.6285
Taxis	0.4628
Minibus	0.5462
Taxis cleanliness	0.4609

Furthermore, the third type of validity, called construct validity, must also be tested. There are two types of construct validity, called convergent and divergent. The divergent construct validity could not be tested because no other construct than quality was measured in the questionaire. To demonstrate convergent construct validity, we have verified that the relationship between the overall quality and the fact that a user has made a complaint or not is negative. Thus, users who have already made a complaint should perceive a lower quality than those who have never complained. According to the t-test results, the average quality perceived by users who have already made a complaint ($n = 77$, average = 7.7) is significantly lower ($p = 0.0008$) than those who never have ($n = 229$, average = 8.5). Furthermore, users who have suffered at least one delay of 30 minutes or more in the past month should perceive a lower quality than the others. According to the t-test results, the average quality perceived by the users who have suffered a delay ($n = 107$, average = 7.4) is significantly lower ($p < 0.0001$) than for those who have never suffered such a delay ($n = 184$, average = 8.7). Again, the theoretical relationship between perceived quality and having suffered a delay is supported by the data. We can therefore conclude that the questionaire is valid according to the convergent construct criterion.

Having tested the reliability, the validity of content, the criterion-related validity and the convergent construct validity of the questionaire, we conclude that it is reliable and valid according to these criteria.

3.3 Importance of the criteria

In the second part of the questionaire we asked the respondents to determine which were the most important criteria to take into account when trying to improve quality of service. The following criteria were among the most frequently mentioned: “if possible, allow last-minute changes in your reservation and allow emergency reservations” (frequency = 55), “always have the same driver for a regular ride” (frequency = 31), “shorten the waiting time for on-call returns” (frequency = 23), and “shorten the time slot to under 30 minutes” (frequency = 20). Most of the measurable criteria were listed among the most important changes. Thus, one could improve quality by incorporating these criteria in the routing and scheduling of the vehicles.

In this section of the questionaire, we also asked the respondents to rank the five measurable quality criteria to take into account in the routing and scheduling of the vehicles: (1) shorten waiting times for on-call returns, (2) shorten the maximum time spent aboard the vehicle, (3) reduce the number of persons that the vehicle can bring back or pick up during your trip, (4) shorten delays in reaching your destination, and (5) shorten the scheduled time slot to under 30 minutes. The ordering depends on the method used to rank the elements. Table 7 presents the differences in the ordering of the five criteria depending on the rules used. We observe that the third criterion is always last and that criteria 1 and 4 are always among the most important. Thus, “shortening the waiting times for on-call returns (1)” and “shortening delays in reaching your destination (4)” seem to be more important than “reducing the number of persons that the vehicle can bring back or pick up during your trip (3)”.

Table 7: Ordering of five service quality criteria according to different ranking methods

Ranking method	**Criteria rank from the most important to the least**
Average	4 - 1 - 5 - 2 - 3
Frequency, most important	1 - 4 and 5 - 2 - 3
Frequency, least important	4 - 2 - 1 - 5 - 3
Frequency weighted sum	1 - 4 - 5 - 2 - 3

These conclusions are useful when trying to redesign a service or to understand the impact on the perception of service when some changes are made in the service offer. However, these conclusions cannot be generalized because the expectations of the users depend to some extent on the current quality of service. Thus, the importance (and expectations) of an attribute are correlated with the perceptions of the quality of this attribute, as observed by Oh (2001). Therefore, these are not static and should be revised periodically. However, our questionaire could prove a useful tool in measuring these indicators periodically.

3.4 Control variables

There exist some significant differences in the perceptions of the overall quality between groups of respondents with respect to age, level of activity, level of schooling, type of handicap, number of years they have used the service, type of ride usually performed (regular or occasional), main reason to use the service and frequency of use (Table 8). Furthermore, users' perceptions of quality do not seem to be different according to the sex of the user, borough of residence, type of vehicle used more frequently by the user, level of revenue and the fact that the user is captive of the service or not.

As concluded by Pagano and McKnight (1983), offering a quality dial-a-ride service is complex because quality depends on many attributes and dimensions, and differents groups of users do not necessarily have the same criteria of quality or the same level of expectations and needs.

4 Conclusions

The aim of this paper was to better understand quality in dial-a-ride services, and to measure this concept by developing a valid and reliable measurement scale. Interviews with dial-a-ride users in a large Canadian city have provided us with the basic elements to construct a measurement scale specific to dial-a-ride services. Fifty-six attributes were identified, compared with the attributes found in the literature, and revised by the service manager and the users association's manager. User perspective was thus considered to measure quality of service. The questionaire developed has proved to be reliable and valid. Our results can help the provider identify the attributes that could be improved and those that should be kept at the same level in order to ensure the desired service level. An exploratory factor analysis of the data collected has allowed us to identify 13 dimensions of quality in dial-a-ride services.

Moreover, this study has identified which criteria are the most important for users and how users rank the most important measurable criteria. This study also identifies control variables by which subgroups of users having different perceptions levels can be categorized. Using these informations, managers should be able to better understand the needs and perceptions of their users and design their service accordingly. They could also use the information on measurable criteria to minimize user inconvenience when optimizing the routing and the scheduling of the vehicles.

Acknowledgements

This work was supported by the Social Sciences and Humanities Research Council of Canada and by the Canadian Natural Sciences and Engineering Research Council under grants 227837-09 and 39682-10. This support is gratefully acknowledged.

Table 8: Comparison of the average scores on the perception of the overall quality of dial-a-ride service across groups formed by control variables.

Control variable	Groups	n	Mean	SD	t-test or ANOVA F-test	p-value
Age					$t(311) = -6.11$	< 0.0001
	[18, 64]	165	7.72	1.98		
	[65, +]	148	8.96	1.54		
Level of activity					$t(307) = -4.90$	< 0.0001
	Active	58	7.28	2.10		
	Inactive	251	8.56	1.72		
Level of schooling					$F(3, 285) = 5.71$	0.0008
	Primary	70	8.54	1.97		
	Secondary	122	8.74	1.54		
	College	51	7.82	2.00		
	University	46	7.67	1.98		
Type of handicap					$F(5, 290) = 4.36$	0.0008
	Ambulatory	80	8.76	1.49		
	Wheelchair	33	8.58	2.12		
	Visual impairment	46	7.74	1.84		
	Intellectual disability	22	8.68	1.29		
	Organic disability	53	7.43	2.03		
	Others	62	8.34	2.20		
Number of years they have used the service					$F(2, 307) = 8.65$	0.0002
	1 year or less	34	9.24	1.02		
	2 to 5 years	134	8.42	1.82		
	More than 5 years	142	7.84	2.07		
Type of ride (regular/occasional)					$F(4, 302) = 2.49$	0.0432
	Occasional, within territory	73	8.74	1.56		
	Occasional, outside territory	8	7.38	3.46		
	Regular, within territory	115	8.04	2.06		
	Regular, outside territory	9	7.33	3.00		
	Others	102	8.28	1.67		
Main reason for using the service					$F(4, 298) = 6.48$	< 0.0001
	Job/school	63	7.29	2.22		
	Leisure and shopping	49	8.16	1.791		
	Medical appointments	131	8.62	1.72		
	Visits to family/friends	27	8.89	1.48		
	Others	33	8.36	1.69		
Frequency of use in the last month					$F(4, 290) = 6.53$	< 0.0001
	0	43	8.60	1.79		
	[1, 10]	163	8.56	1.66		
	[11, 20]	60	7.67	2.02		
	[21, 30]	23	6.83	2.64		
	[31, +]	6	8.66	1.51		
Sex					$t(314) = -1.78$	0.0752
	Male	106	8.01	2.14		
	Female	210	8.41	1.77		
Borough of residence					$t(310) = 0.96$	0.3366
	Longueuil	143	8.36	1.66		
	Others	169	8.15	2.12		
Type of vehicle most often used					$F(3, 305) = 0.90$	0.4422
	Minibus	129	8.20	2.02		
	Accessible taxi	17	8.06	1.89		
	Regular taxi	96	8.09	1.96		
	Others	67	8.57	1.67		
Captive or not					$t(302) = 0.03$	0.9730
	Not captive	119	8.22	1.87		
	Captive	185	8.21	1.97		
Level of revenue					$F(3, 272) = 1.61$	0.1884
	[0, 9999]	98	8.04	1.98		
	[10000, 29999]	134	8.54	1.74		
	[30000, 49999]	29	8.48	1.68		
	[50000, +]	15	8.00	2.00		

References

Abalo J, Varela J, Manzano V (2007) Importance values for importance-performance analysis: A formula for spreading out values derived from preference rankings. Journal of Business Research 60:115–121

Arnould E, Epp A (2006) Deep engagement with consumer experience. In: Grover R, Vriens M (eds) The Handbook of Marketing Research, Sage Publications, London

Bishop G (1987) Experiments with the middle response alternative in survey questions. Public Opinion Quarterly 51:220–232

Brady M, Cronin JJ (2001) Some new thoughts on conceptualizing perceived service quality: A hierarchical approach. Journal of Marketing 65:34–49

Carrillat F, Jaramillo F, Mulki JP (2007) The validity of the SERVQUAL and SERVPERF scales: A meta-analytic view of 17 years of research across five continents. International Journal of Service Industry Management 18:472–490

Cordeau JF, Laporte G (2007) The dial-a-ride problem : Models and algorithms. Annals of Operations Research 153:29–46

Denson CR (2000) Public sector transportation for people with disabilities: A satisfaction survey. Journal of Rehabilitation 66:29–37

Devlin S, Dong H (1994) Service quality from the customers' perspective. Marketing Research 6:4–13

Grönross C (1984) A service quality model and its marketing implications. European Journal of Marketing 18:36–44

Karatepe O, Yavas U, Babakus E (2005) Measuring service quality of banks: Scale development and validation. Journal of Retailing and Consumer Services 12:373–383

Knutsson S (1999) Valuing rider quality attributes in the Swedish special transport services. In: Urban Transport Systems, 2nd KFB-Research Conference, Lund University, Department of Technology and Society, Lunds Tekniska Högskola, Sweden

Lovelock C, Wirtz J, Chew P (2009) Essentials of services marketing. Prentice Hall, Singapore

Martinez Caro L, Garcia JAM (2007) Measuring perceived service quality in urgent transport service. Journal of Retailing and Consumer Services 14:60–72

Neslin S (1981) Linking product features to perceptions: selfstated versus statistically revealed importance weights. Journal of Marketing Research 18:80–86

Oh H (2001) Revisiting importance-performance analysis. Tourism Management 22:617–627

Pagano AM, McKnight CE (1983) Quality of service in special service paratransit: The users' perspective. Transportation Research Record 934:14–23

Paquette J, Cordeau JF, Laporte G (2009) Quality of service in dial-a-ride operations. Computers & Industrial Engineering 56:1721–1734

Parasuraman A, Zeithaml VA, Berry LL (1985) A conceptual model of service quality and its implications for future research. Journal of Marketing 49:41–50

Parasuraman A, Zeithaml VA, Berry LL (1988) SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality. Journal of Retailing 64:12–40

Powpaka S (1996) The role of outcome quality as a determinant of overall service quality in different categories of services industries: An empirical investigation. Journal of Services Marketing 10:5–25

Richards L (2005) Handling qualitative data: A practical guide. Sage Publications, London

Rust RT, Oliver RL (1994) Service quality: Insights and managerial implications from the frontier. In: Rust RT, Oliver RL (eds) Service Quality: New directions in theory and practice, Sage Publications, London, pp 1–19

Sampson S, Showalter M (1999) The performance-importance response function: Observations and implications. Service Industries Journal 19:1–25

Schneider B, White SS (2004) Service Quality, Research Perspectives. Sage Publications, London

Seth N, Deschmukh S, Vrat P (2004) Service quality models: A review. International Journal of Quality & Reliability Management 22:913–949