

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.



Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

**ASSESSMENT OF QUALITY OF MUNICIPAL SERVICES AND
RESIDENTS' SATISFACTION WITH THESE SERVICES**

Claude Roy

A Thesis
in
The Faculty
of
Commerce and Administration

Presented in Partial Fulfilment of the Requirements
for the Degree of Master of Science in Administration at
Concordia University
Montréal, Québec, Canada

July 1997

© Claude Roy, 1997



National Library
of Canada

Acquisitions and
Bibliographic Services

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale
du Canada

Acquisitions et
services bibliographiques

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-40204-5

ABSTRACT

Assessment of Quality of Municipal Services and
Residents' Satisfaction with These Services

Claude Roy

This study examines the adaptation of a Marketing instrument, SERVQUAL (Parasuraman, Berry, & Zeithaml, 1991a), to gauge empirically the attitudes of residents towards the quality of municipal services. SERVQUAL comprises two batteries of items, one for expectations and one for perceptions, and posits that there are five dimensions to the quality construct.

The aims of this study are (a) to develop a reliable and valid instrument to determine the quality of municipal services as perceived by residents, (b) to assess the relative importance of the dimensions in the shaping of the quality construct, (c) to develop a reliable and valid instrument to measure residents' satisfaction with the services provided by a municipal government, and (d) to study the relationship between measures of satisfaction and quality of service offered by that municipality.

A mail survey was conducted on a representative sample of the population of a town in Western Québec. Several findings have emerged from the data analysis. First, the quality construct can be structured hierarchically as a second-order factor model. Second, the importance weights that the residents assign to the dimensions for the expectations are not the same as those for the perceptions. Third, perceptions alone are sufficient to explain the shaping of satisfaction. Hence, this study has successfully resolved some of the outstanding issues concerning the scales, but has raised new issues about their use in the municipal context.

ACKNOWLEDGEMENTS

I begin these acknowledgements by sincerely thanking my supervisor, Professor Jamshid Etezadi-Amoli, Ph.D., for his boundless availability, patience, and dedication while guiding me.

I would also like to extend these thanks to the other members of the thesis committee, Professors Mohan Gopalakrishnan, Ph.D., and Danielle Morin, Ph.D., who appropriately suggested many significant improvements to the investigation and the text.

More thanks go to the staff of the Interlibrary Loans Department of Webster Library who deployed considerable ingenuity and effort to ferret all these documents that were so precious during the elaboration of the literature review.

I am deeply grateful to the Faculty of Commerce and Administration Research Committee for the Master Student-Faculty Thesis Research Grant awarded, to the Fonds commémoratif Louise-Montpetit-Morrisseau of the Syndicat des professionnels et des professionnelles du gouvernement du Québec for their generous scholarship, and to the Chambre de commerce du Haut-Richelieu as well as the Town of Saint-Jean-sur-Richelieu for their cooperation.

Also worthy of my recognition is my friend Carolle, who generously edited my manuscripts.

My sincere gratitude goes to my parents, Germaine and Euclide, whose substantial material contribution was instrumental in the success of this endeavour.

Last but not least, I wish to thank my companion, Gisèle, who affectionately encouraged me at every step of this journey.

TABLE OF CONTENTS

List of Figures	ix
List of Tables	x
List of Appendices	xi
Introduction.....	1
Review of the Relevant Literature.....	4
Determinants of Satisfaction with Municipal Services	4
Socio-Demographic Attributes of the Residents.....	7
Geographic Location of the Residents	15
Political Attitudes of the Residents.....	18
Classes of Services.....	20
Objective Measures of Service.....	22
Subjective Measures of Service.....	25
Concluding on the Different Approaches.....	28
Perceived Quality of Service	28
Distinctive Attributes of Services.....	30
Similarities Between Commercial and Public Services	33
Modelling the Perceived Quality of Service.....	38
Development of the SERVQUAL Scale	42
Conceptual Underpinnings.....	42
Introduction of the SERVQUAL (1988) Scale.....	48
Critiques of the SERVQUAL (1988) Scale	51
Refinement of the SERVQUAL Scale	52
Critiques of the SERVQUAL Scale.....	56
Recent Developments	57
Issues with the SERVQUAL Scale.....	59
Negatively Worded Items	60
Dimensionality.....	60
Universality	62
Relative Importance of the Dimensions.....	63
Expectations.....	64
Difference Scores.....	67

Summary of the Issues with the SERVQUAL Scale.....	77
Applications of the SERVQUAL Scale to Municipal Services.....	77
Crompton and Mackay (1989).....	77
Carlson and Schwarz (1995).....	79
Donnelly et al. (1995).....	80
Summary	80
Measures of Satisfaction and Relationship to Quality.....	81
Measuring Satisfaction with Municipal Services	81
Distinguishing Satisfaction from Quality	84
Linking Quality and Satisfaction.....	87
Summary of the Relevant Research	89
Conceptual Framework and Research Questions.....	90
Objectives of the Research	90
Construct of Quality.....	91
Relative Importance of the Dimensions.....	92
Construct of Satisfaction	93
Relationship Between Quality and Satisfaction	93
Methodology	97
Instrument Design.....	97
Independent Measures	97
Dependent Measures	98
Socio-Demographic Measures	101
Sequencing of the Items.....	103
Packaging	104
Pre-Testing of the Instrument.....	105
Data Sources and Collection	106
Population.....	106
Sampling Design	107
Sampling Plan	108
Collection Procedure.....	109
Statistical Analysis	109
Descriptive Statistics.....	110
Model Estimation.....	111

Psychometric Properties of the Scales	121
Investigation of the Research Questions.....	126
Results and Discussion.....	131
Coding and Validation of the Data	131
Data Entry	131
Data Accuracy.....	132
Anomalous Data.....	132
Sampling Results.....	135
Comparison of the Sample to the Population	135
Early and Late Responses	138
Representativeness of the Sample.....	140
Descriptive Statistics	141
Expectations, Perceptions, Differences Scores	141
Importance Scores.....	144
Satisfaction Scores.....	149
Preparation of the Data for Multivariate Analysis.....	149
Missing Data	149
Appropriateness of the Data Set for Factor Analysis.....	151
Outliers	153
Model Estimation.....	154
Evaluation of Model Identification	154
Model Assessment	154
Re-Specification	158
Psychometric Properties of the Scales	163
Factor Pattern	163
Content Validity	165
Criterion-Related Validity	166
Construct Validity	166
Summary Concerning the Psychometric Properties of the Scales.....	168
Testing of the Research Questions	168
Research Question 1	168
Research Question 2	173
Research Question 3	177

Research Question 4	177
Summary and Conclusion.....	183
Summary of the Findings	183
Implications	184
Practical Implications	184
Theoretical Implications	186
Methodological Implications.....	189
Limitations	190
Prospects for the Future.....	191
Contributions	192
References.....	194

LIST OF FIGURES

Figure 2.1.	Mapping of the literature review	5
Figure 2.2.	Framework for residents' satisfaction with municipal services	8
Figure 2.3.	Service quality model	44
Figure 2.4.	A second-order factor model for quality	95
Figure 2.5.	Path diagram of the simultaneous model.....	96
Figure 3.1.	Illustration of the simultaneous model	117
Figure 4.1.	The transformed quality scale.....	145
Figure 4.2.	Relative importance scores	146
Figure 4.3.	Comparison of the relative importance scores.....	148
Figure 4.4.	Partial standardized estimates of the measurement parameters for the simultaneous model	157
Figure 4.5.	Partial standardized estimates of the structural parameters for the simultaneous model	159
Figure 4.6.	Partial standardized estimates of the structural parameters for the perception-only model	164

LIST OF TABLES

Table 2.1.	Comparison of Different Typologies for Residents' Satisfaction With Municipal Services	9
Table 2.2.	Determinants of Service Quality	47
Table 2.3.	Specifications of the SERVQUAL (1988) Scale.....	50
Table 2.4.	Some Critiques of the SERVQUAL (1988) Scale.....	53
Table 2.5.	Some Critiques of the SERVQUAL Scale.....	58
Table 3.1.	Summary of the Research Questions.....	130
Table 4.1.	Comparison of the Sample's and the Populations' Socio-Demographic Attributes.....	136
Table 4.2.	Batteries and Gap Dimension Scores, Ranking, Comparison with Another Study.....	143
Table 4.3.	Fit Indices for the Alternate Models	155
Table 4.4.	Fit Indices for the Re-Specified Models	160
Table 4.5.	Fit Indices for the Unifactorial Dimensions	165
Table 4.6.	Likelihood Ratio Paired Test Values for Unifactorial Dimensions.....	167
Table 4.7.	Exploratory Factor Analysis Results.....	169
Table 4.8.	Fit Indices for the Basic Structure Models.....	171
Table 4.9.	Fit Indices for the Relative Importance Assessment Models	174
Table 4.10.	Comparison of the Rank-Order of the Relative Importance of the Dimensions.....	175
Table 4.11.	Statistics for the Multigroup Analysis.....	180

LIST OF APPENDICES

Appendix A.	Determinants of Satisfaction with Municipal Services	208
Appendix B.	Survey Material.....	213
Appendix C.	Socio-Demographic Attributes of the Respondents	223
Appendix D.	Summary Descriptive Measures of the Survey Items and Difference Scores.....	226
Appendix E.	Means, Standard Deviations, and Correlations of the Data Set Used for Factorial Analysis	231
Appendix F.	EQS Syntax for the Estimation of the Simultaneous Model.....	238
Appendix G.	Estimates of the Standardized Factor Loadings and Statistics for the Simultaneous Model	244
Appendix H.	Estimates of the Standardized Factor Loadings and Statistics for the Perception-Only Model.....	247
Appendix I.	Estimates of the Standardized Factor Loadings for the Unifactorial Models.....	248

INTRODUCTION

Cities and towns are in turmoil. Several phenomena are perturbing the relatively stable environment that municipal organizations have enjoyed to date. First, the devolution of responsibilities from higher jurisdictions to the local level has “increased the complexity of local governments’ relationships with other levels of government and the private and nonprofit sectors” (Cigler (1993), as cited in Cigler, 1996, p. 60), broadening the mix of services they deliver. Among the new services that towns in the Province of Québec must now assume, many are non-traditional in nature, such as bridge maintenance, which was until recently a provincial responsibility.

Second, complementary to the new-service-without-financing dilemma is the less-financing-for-the-same-service syndrome. An example of this is the recent decision by the Province to discontinue the partial refund of the sales tax expense on municipal purchases. As of 1997, this reduces the revenues of cities by as much as 10 million dollars each, with identical service expenses.

Third, there is a shift in the demographic shape of the population as baby boomers are approaching retirement. In cities, “the fastest growing age group—seniors over seventy-five years old— increase demand for [costly services, such as] . . . transportation services, recreation, and law enforcement” (Cigler, 1996, p. 62). These changes in the social composition of population have resulted in a growing demand for services, not only in terms of the quantity of these services, but also in “increasing scope, range and cost” (Cigler, 1996, p. 65). The author attributed this surge to “a more educated and skeptical citizenry”, increased activism, which manifest itself through zestier interest groups, and technological improvements, with resulting hard-to-resist pressures on elected officials (Cigler, 1996, p. 65-66).

Intuitively, it would seem that these changes, as well as other similar changes yet to come, are likely to affect residents adversely. To monitor the effect of changes in services on residents, an appropriate outcome measure as well as an enabling instrument is required.

Drebin and Brannon (1990) stated that measuring the satisfaction of citizens for services appears to be the *only* effective way to assess if the services that are provided are fulfilling the local community needs. The idea that citizen evaluations are key to assessing the success of municipal endeavours has been advocated for some time (e.g., Stipak, 1974). However, this topic has not been the object of much academic interest (Roth, Bozinoff, & MacIntosh, 1990). Moreover, compared to public services in general, “at a municipal level, there is even less emphasis on measuring consumer satisfaction” (Das, Das, & McKenzie, 1995, p. 77). The search for the appropriate assessment tool is continuing because the outcome measure may still be only weakly defined.

The current study examines the shaping of satisfaction with municipal services in general, and relates it in particular to the quality of the services provided, as perceived by the residents. In addition, this study proposes an instrument to gauge empirically the attitudes of residents towards quality and satisfaction with local services. This instrument is based on a widely used tool to assess the perceived quality of commercial services.

This research is important for municipal deciders in many ways. First, there is presently no framework explaining the relationship between the quality of municipal services and resident satisfaction. Getting a better grasp on what makes the constituents satisfied is capital to public deciders because “presumably one of the main purposes of citizen satisfaction surveys is to provide information for city officials to use in making decisions about service delivery” (Brown & Coulter, 1983, p. 50). Shedding more light on the issue is a valuable endeavour because municipal services affect the entire population and the quality of these services has a direct impact on the quality of life for citizens.

Second, the above mentioned constraints to the delivery of municipal services have begun to affect the level as well as the assortment of services offered. Identifying the

dimensions of quality that citizens prioritize when evaluating services may point to reductions in the level of service quality that do not affect satisfaction because, according to Garvin (1987), “a common error is to introduce dimensions of quality that are unimportant to consumers” (p. 109). These results may also enable the discrimination of areas targeted for improvement on the basis of their significant bearing on satisfaction. This may come handy in an era where “public administrators are feeling pressed to do more with less” (Kelada, 1996, p. 111).

Third, also of considerable value to municipal service providers, this instrument may enhance their current performance indices by the addition of market assessment. In assessing the achievement of their organizational goals, municipal decision-makers are currently restricted to internal measures of service delivery. This study will provide an exogenous and externally focussed managerial feedback on their productivity.

This report begins with a review of the literature concerning satisfaction with municipal services, the assessment of quality with services and the manner in which these attitudes could be linked. The following section examines the methodology used for the study, such as the measures used and the survey sampling process followed. Subsequent sections detail the results, examining the representativeness of the sample, statistics concerning the measures, and the answers to the research questions, as well as cover the implications of the findings, the limitations of this study, and prospective research paths.

REVIEW OF THE RELEVANT LITERATURE

The objectives of this literature review are (a) to set this research in the broader context formed by prior studies, (b) to identify and report the results of closely related studies, and (c) to establish benchmarks to which the current results can be compared (Creswell, 1994).

First, the studies concerned with the shaping of satisfaction with municipal services will be examined. Next, the context of services in general will be considered, along with an assessment of the similarities between commercial services and local¹ services, as well as an examination of how the perceived quality of service has been defined. Then, a particular attention will be dedicated to an instrument used for the measurement of quality with commercial services, namely SERVQUAL (Parasuraman, Zeithaml, & Berry, 1988) and to its use with municipal services. Last, the topic of the measurement of satisfaction will be examined. The mapping of the literature review is illustrated in Figure 2.1.

Determinants of Satisfaction with Municipal Services

In this sub-section, studies that have a focus on explaining possible causes of satisfaction with municipal services will be examined. Attempts to explain the satisfaction of residents with municipal services have resulted in a prolific body of knowledge, with considerable devoted effort and ingenuity. In their quest, authors have sought to establish a relationship between satisfaction, as an outcome variable, and some explanatory variables. This line of investigation aimed principally at demonstrating some uniformity in satisfaction among groups, with noticeable distinctions between groups. For example, men could be more satisfied than women could be with some local services.

¹ The terms *municipal* and *local* will be used indiscriminately when referring to services provided by a lower-level jurisdiction to its residents, as both terms are often used as synonyms in the literature.

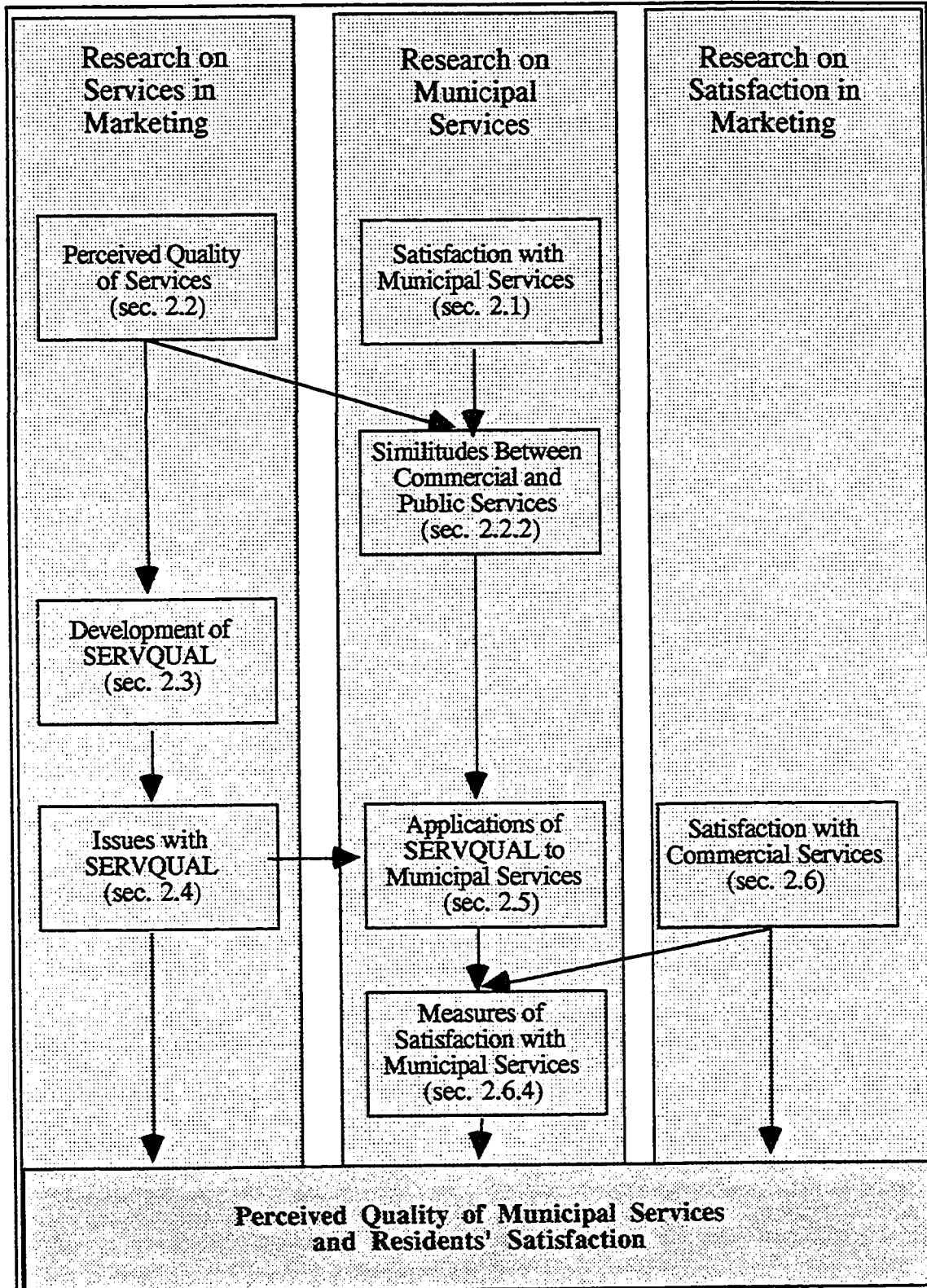


Figure 2.1. Mapping of the literature review.

Many factors have been investigated for their conceptual and empirical contributions to the shaping of satisfaction with municipal services. The authors of some studies examined a wide array of variables. However, most focussed on a limited set of variables that shared some similarities, such as demographic attributes of the residents or different objective measures of service delivery.

Three groups of authors have proposed typologies to classify these explanatory variables of interest into a structured framework. Examining only citizen-based measures, Brudney and England (1982) used a four-class typology to group determinants of satisfaction with municipal services. These classes are (a) spatial characteristics, such as neighbourhood attributes; (b) personal attributes, such as age or race; (c) political attitudes; and (d) housing characteristics, like home ownership.

Under Taylor's (1982) classification, the determinants also fall into four categories: (a) group socio-demographics, which are similar to Brudney and England's personal attributes; (b) service, meaning agency-based objective data; (c) attitudes, mainly political; and (d) area socio-demographics, concerned with the specifics of the delivery area.

Lyons, Lowery, and DeHoog (1992) proposed a three-level grouping, with overlapping variables. These groups are (a) individual-level explanations, including demographics, home ownership and attitudes; (b) jurisdiction-level explanations, including some demographics, community expectations, governance structure and objective measures; and (c) neighbourhood-level explanations, measured arbitrarily through dummy variables attributed to the different sub-areas forming the wider investigation area.

An examination of the relevant literature, which is detailed in the following subsection, revealed that other potential determinants of satisfaction cannot be classified under the above typologies. In order to accommodate for these potentially significant determinants, the current study builds on the previously mentioned frameworks and proposes a new, encompassing typology.

In this new framework, the variables of interest for explaining satisfaction with municipal services are gathered into six mutually exclusive and homogeneous clusters of determinants, labeled as *classes*. These six classes form two sub-sets: the attributes of the residents and the attributes of the service itself. For the residents' attributes, the classes are (a) the *socio-demographic attributes* of the residents, (b) the *geographic location* of the residents, and (c) the *political attitudes* of the residents. For the attributes of the service itself, the classes are (d) the *classes of services*, which emphasizes similarities in the intrinsic nature of the services; (e) the *objective measures* of service; and (f) the *subjective measures* of service. The framework is illustrated in Figure 2.2.

As can be observed from Table 2.1, elements of the three previously reported classifications generally fall within four of the six proposed classes. The two additional classes are the classes of services, which received little scholarly interest previous to the 1990s and which have since then been investigated primarily in studies performed in Canada, and the subjective measures of service, an avenue which has only recently been endowed with tools appropriate to the rigorous investigation of attitudes. A review of pertinent studies for each class follows. Most of these studies refer to settings in the United States of America.

Socio-Demographic Attributes of the Residents

The authors of studies examining variables that fall under the socio-demographic attributes of the residents' class examined the idea that individual characteristics of the residents, such as age or gender, are the determinants of satisfaction with municipal services. At the outset, this approach appears very intuitive. Numerous researchers in a wide array of disciplines have indeed followed the path of individual differences. Moreover, many of these studies did reveal significant differences on the basis of some personal attribute. Not surprisingly, this tradition has caught on early in the investigation of the determinants of satisfaction with municipal services and has resulted in a rich body of literature.

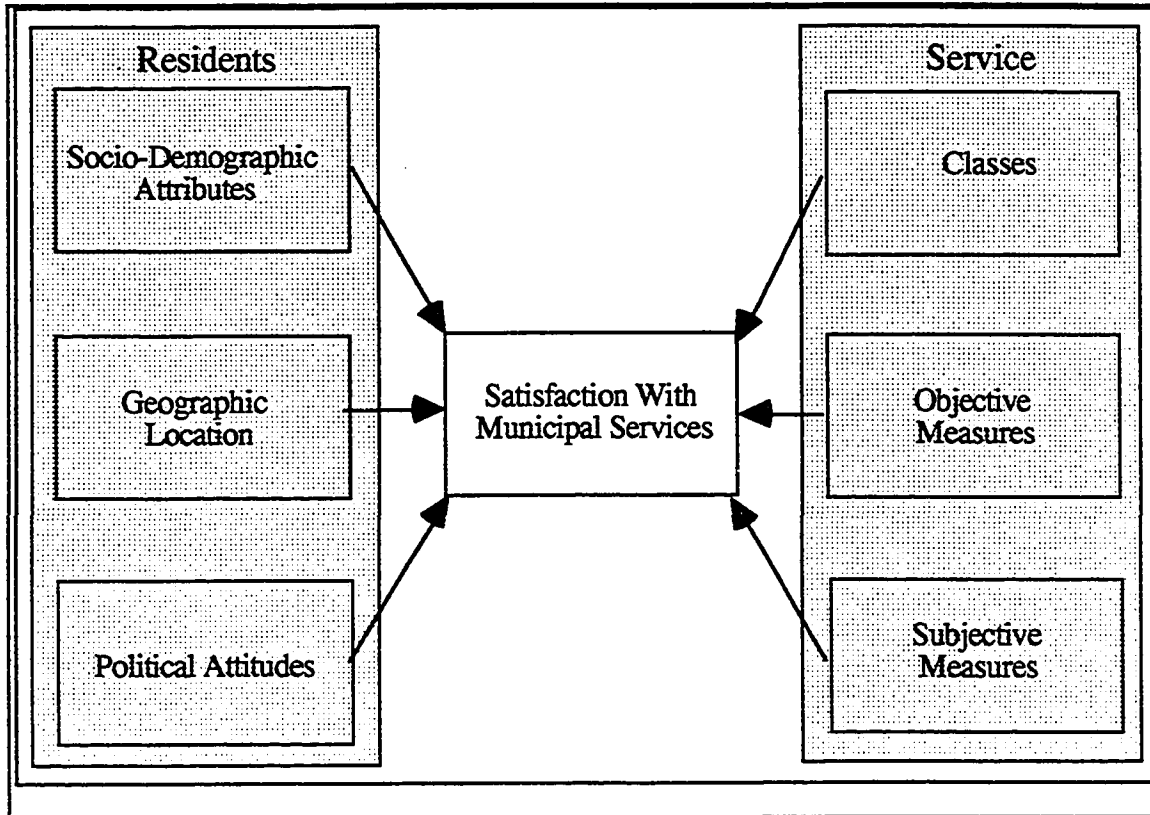


Figure 2.2. Framework for residents' satisfaction with municipal services.

Table 2.1.

Comparison of Different Typologies for Residents' Satisfaction With Municipal Services

Brudney & England (1982)	Taylor (1982)	Lyons et al. (1992)	Current study
Personal attributes, housing characteristics	Group socio- demographic attributes	Individual-level explanations, jurisdictional-level explanations	Socio-demographic attributes of residents
Spatial characteristics, housing characteristics	Area socio- demographic attributes	Neighbourhood- level explanations.	Geographic location of residents
Political attitude	Attitudinal	Jurisdictional-level explanations	Political attitudes of residents
--	--	--	Classes of services
--	Service	Jurisdictional-level explanations	Objective measures of service
--	--	--	Subjective measures of service

Note. Dashes (--) indicate that the authors did not examine this category of attributes.

The variables of interest under the socio-demographic attributes of the residents' category are (a) race, (b) age, (c) gender, (d) income, (e) education, (f) social class status, (g) occupation, and (h) home ownership.

Race

Race was often investigated as a predictor of satisfaction with municipal services. Some studies examining the influence of race indicated that Afro-Americans were generally less satisfied with municipal services than Caucasians, with regards to police services (Brown & Coulter, 1983) and general services (Lovrich & Taylor, 1976; Stipak, 1974). The race cleavage phenomenon was also observed for persons of Mexican descent with regards to general services (Lovrich & Taylor, 1976), as well as garbage collection, and park and recreation (Stipak, 1974). In addition, the same cleavage was observed for Native Americans when examining services in general (Pelissero, 1978).

However, the race difference was not observed for services in general between Blacks and Hispanics in Oklahoma, an area that is mainly rural (Pelissero, 1978). This finding contrasts with the previously reported studies that were conducted predominantly in urban areas. Pushing the analysis one step further, Lovrich and Taylor (1976) found that race remained a critical factor in explaining satisfaction, even when the prevalent socio-economic differences between races were controlled for.

There is evidence, however, to support an opposing view that ethnic differences appear significant *only* because other variables that may share collinear relationships, such as income or alienation, have not been properly controlled for. Tan and Murrell (1984) found that race was only important for satisfaction with police services, and then only in conjunction with a generally negative attitude towards the community. Considering satisfaction with police services in military communities, Brede (1985) found that the effect of race was negligible when other variables were accounted for. Similar non-significant findings were reported in a study that examined racial differences while controlling for autonomy of the delivery agent (Rosentraub & Thompson, 1981).

In summary, the influence of race on satisfaction seems doubtful. This lack of evidence confirms the proposition that "blacks seem no more ideological in their attitudes toward city services than did whites" (Brudney & England, 1982, p. 130).

Age

Age has also been considered instrumental in explaining the satisfaction level of residents. In a study on satisfaction with community services, which included municipal services as well as other services, such as shopping, "the main effects are also significant for age" (Murdock & Schriener, 1979, p. 117). Age was also found to be positively and significantly related to satisfaction with police services (Brown & Coulter, 1983; Stipak, 1974) and police response time (Percy, 1986). The same relationship was observed for generalized services, for garbage collection, and for park and recreation, but not for other services (Stipak, 1974).

However, no difference between satisfied and dissatisfied customers on the basis of age was found in a study on police telephone report units (Glauser & Tullar, 1985). Similarly, no relationship was observed between age and satisfaction in a study on several municipal services (Das et al., 1995).

In a study specifically focussed on the relationship between age and satisfaction for welfare recipients², Goodsell (1981) found that “older clients tend to be more content (or less discontent) with the treatment they receive from the bureaucrats” (p. 2). He also found that the proportion of very satisfied respondents rise in successive age cohorts, ranging from 0% for teenagers to 73% for persons in their sixties or older. The author attributed this propensity of older clients to be more satisfied to positive discrimination in favour of the elderly, resulting in better treatment for the aged, and a generalized more positive assessment by the elderly of the bureaucratic treatment to which they are subjected, because they are more accepting of their fate (Goodsell, 1981).

In summary, although the evidence on age as a determinant of satisfaction is mixed, there is a conceptual motive and some empirical evidence to believe that there is an age bias in the satisfaction with personally delivered local services. However, such services do not represent a large portion of the municipal service basket. Indeed, many municipal services, by nature, are delivered in an indiscriminate fashion. Therefore, the observed bias may not prevail for the bulk of municipal services.

Gender

Differences according to gender have also been investigated. Brown and Coulter (1983) found that “males tend to be slightly more satisfied than females” (p. 54) with police services, while another study on the same topic led to mixed evidence (Stipak, 1974). No difference between satisfied and dissatisfied customers on the basis of gender was found in a study on police telephone report units (Glauser & Tullar, 1985). This absence of relationship was also observed for generalized services (Lyons et al., 1992) and several

² While welfare is primarily a provincial matter, it may nonetheless be considered a local service. For instance, the City for Montréal does the administration of the service for its residents.

services (Roth et al., 1990). In summary, the evidence on gender as a determinant of satisfaction is weak.

Income

Income has been reported to have a significant positive link with satisfaction for police services (Brown & Coulter, 1983), or not to be significant (Stipak, 1974). However, family income differentials were not found to be a significant predictor of satisfaction for overall services (Lovrich & Taylor, 1976; Rosentraub & Thompson, 1981; Stipak, 1974). In the Canadian setting, one study revealed a positive relationship between income and satisfaction for fire department services, but not for other services (Roth et al., 1990), while no relationship was observed for several municipal services (Das et al., 1995). Thus, there is mixed evidence concerning income as a determinant of satisfaction with municipal services.

Education

Education did not reveal to be a significant predictor of satisfaction for police services (Brown & Coulter, 1983; Stipak, 1974), for police telephone report units (Glauser & Tullar, 1985), nor for services in general (Stipak, 1974). This absence of a relationship was also evident in Canadian studies on several municipal services (Das et al., 1995; Roth et al., 1990). In summary, the evidence appears to be against education as a determinant of satisfaction.

Social Class Status

When military rank was employed as an indicator of social class achievement, a cleavage occurred, with "senior-ranking soldiers tending to be more satisfied with military police service" (Brede, 1985, p. 114). However, the author considered that this finding had little external validity. He attributed the relationship to the exceptional "spatially discrete, rank-oriented residential" arrangement of Army communities, where police services may be subject to "allocative choices favoring some sub-populations", rather than proof that "police services are distributed based on indefensible criteria such as economic

advantage and political power” (Brede, 1985, p. 114). In summary, the evidence for social class status as a motivator for satisfaction with local services was assessed in a single study and the author himself suggested that the results be interpreted with caution.

Occupation

Occupation was also tested for its bearing on satisfaction. According to Taylor (1982), respondents with a manual occupation, such as farmers and craftsmen, expressed more satisfaction for services provided by fire prevention departments, library and parks than those with a non-manual occupation, such as clerical or clergy. However, the manual/non-manual class distinction was insignificant for police services. In the Canadian setting, no difference in the satisfaction level was found on the basis of occupation (Roth et al., 1990). In summary, the evidence on occupation as the explaining factor for satisfaction is mixed.

Homeownership

Homeowners displayed higher satisfaction levels for a municipal building inspection service (Stipak, 1974). However, owning a home and having dealings with the municipal building inspection service are not independent events, so their relationship may be more complex than the *prima facie* cause to effect pattern suggested. In another study, the influence of home ownership was insignificant for generalized services (Lyons et al., 1992). Consequently, the evidence of home ownership as an explaining factor for satisfaction is mixed, at best.

Longitudinal Design

In a longitudinal study on police services, Tan and Murrell (1984) found that the explaining variables would lose their significance when the samples were pooled across several years, even when cross-sectional data analysis had revealed that a respondent characteristic was a significant predictor of satisfaction, except for age that remained significant across time periods.

Summary of Socio-Demographic Variables

In general, this intuitively appealing approach has been deceiving. Few of the repeated attempts to link satisfaction with demographic correlates have produced better than mixed evidence. Stronger results were attenuated by doubts concerning the opportunity to extend them to the general population or to all services. In addition, significant findings often vanished when the interaction effects were controlled for, hinting that the covariate may be the instrumental factor.

Even when significant relationships were established, the utility of the findings may be limited. Indeed, some evidence may have been found that race, “age, income, gender, occupation, home ownership, marital status, and size of household may affect some services evaluations, but . . . there is little consensus about the importance of these variables” (Lyons et al., 1992, p. 19), because most studies reporting significant results failed the replication test. Except for age, explanations as to *exactly why* this relationship exists are not available, and these relationships have not been framed through a theoretical underpinning. The data may exhibit correlation, but it is likely that the suspected causal link “remains simply an unexplained empirical observation” (Lyons et al., 1992, p. 19).

Studies that examined interaction effects have raised the possibility that factorial models could uncover significant differences between different segments of the population without producing substantial evidence in favour of this hypothesis. In addition, as the number of segmenting factors increases, the size of the subset decreases, and even if a sizeable difference was tracked to a highly defined group, this finding would lack practical significance in view of the relatively small target sub-population. Therefore, in spite of its appeal, the socio-demographic attributes’ approach failed to deliver a theoretically grounded, empirically verified motivation for variations in the level of satisfaction with municipal services.

Geographic Location of the Residents

Distinctions on the satisfaction level with municipal services may be observed between the communities. Also, communities can be very different in terms of living conditions, ranging from slums to posh neighbourhoods. Building on these two sets of differences, the conceptual root of the current approach is that the conditions prevailing in the community are generally reflected onto satisfaction. The rationale for this line of investigation is that "citizens living in neighborhoods with dilapidated housing may tend to generalize their dissatisfaction with housing conditions to all aspects of the neighborhood, including local governmental services" (Stipak, 1979a, p. 434). Thus, the use of variables from the geographic location of residents' category tests the hypothesis that the physical environment exercises a significant influence on the satisfaction of residents with local government services. The focus is on the following variables (a) neighbourhoods, (b) stage of community development, (c) type of dwelling, and (d) size of jurisdiction.

Neighbourhoods

Neighbourhoods, and notably the differences between them, have been examined in this view. Schuman and Gruenberg (1972) found that "persons living in largely black and lower income areas are most dissatisfied with the services they receive—regardless of their race or income; persons living in largely white and upper income areas are most satisfied with their services—again regardless of their own race or income" (p. 387). In addition to its value concerning the current approach, this study reinforces the evidence reported in the preceding sub-section against race and income as determinants of satisfaction. Pelissero (1978) found that residents of better kept neighbourhoods were more satisfied, while those of poorly kept areas were less satisfied. In addition, Lovrich and Taylor (1976) found that "a very strong relationship [existed] between condition of neighborhood and evaluation of city services generally, and city services in one's neighborhood in particular" (p. 208).

The evidence for this proposition appears compelling, but an alternative explanation can be developed. Differences in satisfaction between neighbourhoods may predominantly

reflect asymmetry in the delivery of services prevailing in the areas, on the basis of their conditions. That is, the conditions prevailing in a particular area would influence the service delivery in that area. Run down boroughs would *need*, *deserve*, and *get* less services on the basis of some area's attributes distribution equity. Thus, it would be these differences in objective measures of service delivery between the areas that would influence satisfaction.

Along this view, it has been asserted that "services are indeed better in suburbs than in central cities, and that this fact is the source of the difference in satisfaction levels" (Schuman & Gruenberg, 1972, p. 377). However, these findings must be weighted against forthcoming evidence concerning the strength of the influence of objective measures of service delivery on satisfaction, which posits that objective measures have only a weak influence on satisfaction. In summary, there is weak evidence that neighbourhoods' conditions may influence satisfaction with local services.

Stage of Community Development

Murdock and Schriener (1979) examined community development's influence on satisfaction by scrutinizing a group of communities from the perspective of their social and economic growth. The authors discovered that residents of communities at the *currently developing stage* of the community life cycle were less satisfied with community services than those in either the *predevelopment* or the *postdevelopment* stages.

Type of Dwelling

Type of dwelling has been hypothesized as an explaining variable for satisfaction. Pelissero (1978) found that residents of single-family dwellings appeared more satisfied with municipal services than residents of multiple-family units.

Size of Jurisdiction

The size of the jurisdiction has been put forward as a possible motivator for satisfaction. Stipak (1974) proposed that smaller political jurisdictions enhance resident satisfaction, by providing an array of services that would be more palatable to residents.

Indeed, in a study examining services in general, as well as parks and recreation, Stipak (1974) found that the increasing size of the local government had a negative impact on satisfaction, such that residents of smaller cities were more satisfied with services than those of larger cities.

However, this contradicts previous evidence that satisfaction for police, garbage and park services was lower in smaller towns (Schuman & Gruenberg, 1972). Similar findings were reported in the Canadian context, as "people in larger communities are more satisfied with fire prevention, garbage collection, and police departments" (Roth et al., 1990, p. 582). The latter interpreted these findings as evidence that municipal services are better in larger centres, more diverse and more available. However, the validity of this argument needs to be assessed in view of the findings for the objective measures' approach, which are forthcoming.

So, while the evidence suggests that the size of the community may be significant, the direction of the relationship is undetermined. It may also be that this presumed relationship is not linear, adopting instead the shape of an hyperbola. The interpretation of this alternatively shaped relationship would be that there is a decrease in satisfaction as the size of the jurisdiction rises from small to medium, with a subsequent increase as the jurisdiction grows larger. Although conceptually different, the shape of the relationship would be similar for the previously mentioned stage of community development variable.

Summary of Geographic Location Variables

In general, here again the approach stems from propositions that are well rooted in the folklore, such as the small town advantage. While the evidence for the geographic location approach is not trivial, it suffers considerable adversary findings. The few undisputed evidence stem from un-replicated studies, rather old, with limited potential for policy implementation in mature jurisdictions.

Political Attitudes of the Residents

There is evidence that residents' mood has a negative impact on their satisfaction with local services (Percy, 1986). In psychology and consumer behaviour, it is a well-accepted proposition that "negative moods are likely to adversely affect the evaluations" (Manrai, 1993, p. 165). However, moods are generally mild, situation specific and transient conditions (Manrai, 1993), which reduce their utility as determinants of satisfaction with municipal services. Consequently, more attention concerning attitudes has been devoted to political attitudinal variables, which demonstrate a better stability and persistence over time.

Various residents' political attitudes have been used in attempts to explain satisfaction. The underlying assumption for the examination of political attitudinal variables is that the attitude towards public services may reflect phenomena occurring at a more general political level. Differences in satisfaction for services may thus principally reflect satisfaction with other political dimensions. The variables of interest under this category are (a) social investment, (b) general assessment of the local government, (c) political efficacy, and (d) interest in government and public affairs.

Social Investment

Social investment, the degree of social and psychological involvement and attachment to the community, has been reported to demonstrate a significant and positive link to satisfaction. High commitment and loyalty to the community have been linked indirectly to a positive evaluation of local services (Lyons et al., 1992). In addition, the same authors obtained evidence that citizens who knitted close social and psychological *liens* in their community were more satisfied. So, social investment demonstrates some potential as a determinant of satisfaction with local services.

General Assessment of the Local Government

Residents who generally assessed their local government more favourably tended to be more satisfied with police services (Brown & Coulter, 1983). Thus, the general

assessment of the municipal government demonstrates some potential as a determinant of satisfaction with local services. However, there is a possibility that the causal link be in the opposite direction. Indeed, it is possible that the aggregation of positive ratings towards individual services lead to a global favourable assessment towards the civic body, not the opposite.

Political Efficacy

Political efficacy is the belief that one is able to affect a government body (Stipak, 1974). A politically efficacious person is strongly convinced that elected officials act responsively and that her/his vote counts. Local political efficacy extends this definition to particular local governments (Lyons et al., 1992). More local politically efficacious residents evaluated local government services positively and were found to be more satisfied with the services they received in general (Lyons et al., 1992; Stipak, 1974) and for police services (Brown & Coulter, 1983). In summary, evidence was found in favour of political efficacy as a significant predictor of satisfaction with local services.

Interest in Government and Public Affairs

Interest in government and public affairs was also tested as a predictor for satisfaction. It revealed to have a negative impact for services in general (Stipak, 1974). These findings are paradoxical when paralleled with those concerning political efficacy. The effect of a strong belief that elected officials act responsively and that one's vote counts should converge with those of interest in government and public affairs. Yet, political efficacy and interest in civic affairs seem to influence satisfaction in an opposite manner.

Summary of Political Attitudinal Variables

The evidence uncovered within this approach appears stronger. In addition, the theoretical underpinnings of the relationship between the political attitudinal variables and satisfaction appear natural. Indeed, all these constructs are defined within the field of political activities or outcomes, a restricted domain. However, in spite of this conceptual proximity, the evidence is again conflicting on the nature of the relationship. Approaches

that examine the satisfaction issue from a perspective that it is inherent to the service itself will be examined in the following sub-section.

Classes of Services

The classes of services approach aims at explaining satisfaction strictly with intrinsic properties of the services, independently from delivery outcomes. The hypothesis is that services of the same type, formed into classes, will generate similar patterns of satisfaction. This line of investigation has been pursued mostly in the Canadian context. The principal variables that have been investigated using this classification method are (a) salience, (b) experience and homogeneity, and (c) discretion available to the delivery agent during the performance of the service.

Salience

In a study encompassing public services emanating from local, provincial and federal levels, Roth et al. (1990) found that "the lower the direct experience with a service[,] the higher proportion of the population is satisfied with a service" (p. 578). The apparent higher levels of satisfaction with services with which the citizens are less familiar can be deceiving. Citizens are requested and willing to express opinions on public affairs, in spite of their relative ignorance of the matter in question, which precludes them from forming "informed or meaningful answers" (Hatry & Blair, 1976, p. 315).

A study by Bishop, Oldenick, Tuchfarber and Bennett (1980) revealed that as much as one third of the respondents had an opinion on a non-existent issue. They posited that such respondents attempted to save face by doing so, as "it does, after all, take a certain amount of confidence in oneself to acknowledge that one does not have an opinion on something that sounds important" (Bishop et al., 1980, p. 206). Thus, the observed high satisfaction levels may simply reflect the fact that "citizens . . . quite willingly provide evaluations of specific local services, despite a lack of knowledge or perceptions of service quality" (Stipak, 1979a, p. 422). High satisfaction with services that are not salient to the citizen would thus be motivated by the absence of knowledge, while low satisfaction with

services which are more familiar to the citizen would then rest on some other evaluation basis.

Another explanation for the inverse relationship between salience and satisfaction may be attributed to a bureaucratic propensity to minimize contacts with the citizen. Referring to federal level public services, Picherak (1987) mentioned that “increasing the frequency of contact reduces the likelihood that the client will be satisfied ultimately in his or her impression of the quality of service. Unnecessary contact produces opportunities for significant misunderstanding that can destroy any favourable impression of the department” (p. 252). This was confirmed in the related field of health care, where “the greater the number of people with whom customers interact during service transactions the less likely customers will report satisfaction” (Whitman-Smithe, 1995, p. 57). Consequently, the influence of salience of the services on satisfaction may not be material.

Experience and Homogeneity

Building on Roth et al.'s (1990) work, Das et al. (1995) classified municipal services on two axis, high experience/low experience—used often or not—and homogeneous/heterogeneous --uniform or adapted provision. This two by two matrix grid was used to interpret a ranking of services according to the magnitude of their satisfaction measures. Das et al. (1995) found that residents were more satisfied with low experience, homogeneous services, such as those provided by the fire department. The authors interpreted that dissatisfaction would significantly increase from “attempt[s] to meet heterogeneous needs of consumers with undifferentiated products and service levels” (Das et al., 1995, p. 90). However, their conclusion could alternately be interpreted to mean that regardless of the type of service, the customer is not satisfied when he/she does not receive the service sought, or when the service is provided in an awkward manner. If such is the case, the classes of services issue is subordinated to the provision of the proper service or to the adequacy of the provision mechanism. Hence, this classification of services does not appear to provide substantial information on the source of the residents' satisfaction.

Discretion in the Performance of the Service Activity

Rosentraub and Thompson (1981) classified the services on their potential for discretionary activities—the ability of the service agent to modulate service distribution patterns. The classes were (a) high, a level typically found in police (Galloway, 1992) and emergency services; (b) medium, such as parks and libraries; and (c) low, such as garbage collection and utilities, which are delivered in a controlled, uniform manner, according to standards, and where the autonomy of the agent is minimal. Rosentraub and Thompson (1981) found that there was more variability in the outcome of high potential for discretion services than in other classes, but the direction of the relationship was not ascertained. Thus, increasing levels of discretion available to the agent might produce more variability in the outcome, but it is not clear how this variable influences satisfaction.

Summary of Classes of Services Variables

In essence, although examining the intrinsic attributes of the services as potential determinants for satisfaction seemed a good path to follow, no sizeable conclusive evidence was found under this approach.

Objective Measures of Service

Within the public services context, objective measures of service delivery are defined, in a strict sense, as the “performance characteristics [of these services] using data from official archives of public agencies” (Brown & Coulter, 1983, p. 50). The objective measures of service encompass “objective indices of performance and results: frequency of garbage collection, low crime and high arrest rates, reading levels of school children” (Schuman & Gruenberg, 1972, p. 369). The standing hypothesis is that the relation with satisfaction is linear and positive. The principal objective measures that have been investigated under this caption are usage and performance measures, often thought of as process quality measures.

Usage

Usage of service has been tested as the determinant of satisfaction in some studies. In favour of a significant relationship, Lyons et al. (1992) found that for generalized services, satisfaction was positively responsive to the number of services provided, in spite of the fact that the cardinal measurement did not account for the differences in services levels—bare-basic to full-fledge service. Using expenditures as a “useful—albeit crude and incomplete—index of service level”, Stipak (1974, p. 43) found only a modest impact between objective measures and satisfaction across an array of services, except for park and recreation services, where a significant relationship existed. Regarding the use of expenditures as a measure of quality, Parks (1984) pointed out that expenditures are a measure of input, not outcome, and that there is a low congruence between the concept and its indicator.

To supplement *hard* data, self-reported factual data on usage have been used as proxies to objective measures, mildly relaxing the above-mentioned definition of objective measures of service delivery. For police services, one study reported a negative relationship between the self-reported number of police contacts and satisfaction (Pelissero, 1978). However, in a study by Brown and Coulter (1983), where frequency was measured distinctly for experience—police encounters—and victimization—being the victim of a crime, the relationships were not as strong. A significant negative relationship still existed between victimization and satisfaction for police services, but experience and satisfaction were no longer significantly related. In summary, the evidence concerning usage as a determinant of satisfaction is mixed.

Performance

Performance measures of service typically include variables such as response time, crime rates, and arrests ratios for police services. Satisfaction was found to be independent from performance for police services (Brown & Coulter, 1983), military police (Brede, 1985), and also for mass transit (Tan & Murrell, 1984).

Discussion of Objective Measures of Service Variables

In summary, findings concerning objective measures of service as determinants of satisfaction lack congruence. The first explanation for this lack of congruence is statistical complications arising from the use of aggregate measures. Jurisdiction-wide averages which have often been used, in view of the absence of more closely related data, lack the variability to account for area-restricted phenomena (Parks, 1984). This contrasts with the citizen's evaluation that was hypothesized to be a comprehensive measure of local conditions in a clearly defined area.

Second, from a conceptual perspective, Stipak (1979b) proposed that the reason why the actual service performance exercised such little influence on satisfaction is that "citizens pay little attention to those services, as long as service quality is within some adequate range" (p. 48). Consequently, contextual salience of the service would arise only from conspicuously extreme levels of performance, in which case some effect on satisfaction scores can be anticipated.

Third, objective measures that are contextual in nature and determined by aggregation of official records, such as crime rate and response time, may not be conspicuous to residents. These measures are likely to have less impact on attitude formation than personal experience. On that basis, Percy (1986) affirmed that the lack of correlation between the official quality measures and citizen satisfaction does not *alone* constitute a sufficient motive to disregard citizen evaluations.

Fourth, the objective indicators of quality may fail to accurately account for satisfaction due to a lack of content validity. Measures may concentrate on a specific aspect of the service and fail to represent a significant portion of the task. For instance, in a cross-section of studies on police services, Parks (1984) found that the indicators used were mostly crime-related, representing sometimes as little as 10% of the overall police activity. He added that the citizen's evaluations might however reflect not only this 10% of the workload, but also the untapped 90%.

In view of these poor results, some researchers have investigated evidence of coalescence between objective and subjective measures. The goal was to assess if citizens were getting substantive information from which to infer their satisfaction evaluations. For street conditions, Carroll (1978) found a significant relationship between objective measures of quality of the service and citizens perceptions of the level of quality. Thus, citizens can tell the difference between a street that has potholes and one that does not. This ability of citizens to properly assess the quality of service was later confirmed for services in general (Rosentraub & Thompson, 1981) and for police response time (Percy, 1986).

Consequently, the absence of a relationship between objective measures of service and satisfaction cannot be attributed to the lack of ability of citizens to form attitudes about the object of focus. Brudney and England (1982) examined previous studies on the matter. They noticed that “the tenuous relationship between subjective and objective indicators may be a function of the type and quality of objective output measures employed in research” and that “research attention should be directed toward obtaining more and better direct observational data on the quality of city services” (p. 129).

In conclusion, objective measures of service did not prove to be worthy antecedents of satisfaction. To explain this, Brown and Coulter (1983) posited that “perhaps one important reason why objective service conditions do not affect satisfaction levels is that citizens interpret objective service conditions through their subjective service expectations” (p. 57). This proposition will be investigated in the next section on subjective measures of service.

Subjective Measures of Service

The subjective measures of service are defined as “attitudes about a service through some form of survey research” (Brown & Coulter, 1983, p. 50). The following variables have been examined: (a) safety, (b) promptness, (c) availability, and (d) quality.

Self-reported factual measures, which were classified as objective measures, could also fall under the present caption. For the purpose of this research, previous studies based

on agency and survey data concerning events were classified as objective, while those focussing on attitudes were labelled as subjective.

Safety

Safety has been tested as a predictor of satisfaction, especially for police services. The feeling of safety while walking in the neighbourhood at night revealed to be a significant predictor of satisfaction (Brown & Coulter, 1983; Pelissero, 1978). McClendon and O'Brien (1988) found that safety was consistently significant for Caucasians. When controlling for income, safety was significant for Afro-Americans only if they resided in a *some crime* neighbourhood and if the items on service indicators preceded the satisfaction items in the survey.

Pelissero (1978) found that the perceived rate of vandalism had a negative impact on satisfaction. Similarly, Percy (1986) found that resident who "saw neighborhood crimes rates as rising tended to evaluate police services more negatively" (p. 81). However, the link between the concept of safety and its measures may be weak on content validity. Surrogate measures may encompass phenomena that are external to the construct. Referring to the above studies, the safety measures may assess more than strictly police services. The feeling of safeness in the streets at night may also be a function of street lightning and the type of neighbourhood. The latter phenomenon could also be instrumental in the perceived rate of vandalism. In addition, individual differences in anxiety, self-confidence or the level of proficiency at martial arts could be responsible for some of the variation in the response. Furthermore, neighbourhood crime rates are partially determined by factors that are outside of police reach, such as poverty. In summary, the evidence on safety as a determinant of satisfaction appears conclusive but weak.

Promptness

Promptness has also been examined for its effect on satisfaction. The perception of delay in response time has been documented as a valid predictor of satisfaction for police services "in particular incidents, and in . . . neighborhoods more generally" (Parks, 1984,

p. 120). Comparing self-reported to agency data, Percy (1986) found that perceptions that “response time was either faster or slower than expected were positively and significantly related to satisfaction for the service” (p. 77). Thus, the impression of promptness appears to be a consistent and significant predictor of satisfaction.

Availability

Availability of the services, measured from the recipients’ perspective, was found to be “most relevant to an understanding of service satisfaction” (Taylor, 1982, p. 107). However, this determinant is likely to be affected by the influence of covariates, in a manner similar to usage.

Quality

The perceived quality of the services was also used to predict satisfaction. Lyons et al. (1992) compared services between jurisdictions using a perceptual measure of service quality and found that perceived service quality revealed to be a significant predictor of satisfaction. In addition, evidence to support the assertion that “a poor service, if seldom experienced, in the aggregate, may be perceived by most consumers as adequate” was found by Das et al. (1995, p. 81). In summary, the evidence on quality as a determinant of satisfaction with municipal services is promising.

Summary of Subjective Measures of Service Variables

The use of subjective measures may be prone to some technical difficulties. However, the bulk of the evidence suggests that subjective measures in general, and perceived quality in particular, can be considered as having a reasonable potential as an antecedent of satisfaction. Expectations may have to be controlled for. A similar view is shared by Pollitt and Bouckaert (1995, p. 17) who posited that “*satisfaction* (or dissatisfaction) is the result of the confrontation of expectations (individual or collective) and perceived quality.”

Concluding on the Different Approaches

Summary findings for the above mentioned studies are profiled in Appendix A. For most determinants, the results were either non significant, mixed, or conflicting as to the nature of the relationship. Only the perceived measures have yielded consistent findings.

The individual methodologies used in these studies were not always reported. When they were, the authors generally indicated utilizing statistical methods that do not accommodate for error in measurements, such as the analysis of variance and linear regression. These techniques produce unbiased estimates of the population parameters *only* when the assumption that all measures are free from error is met. Otherwise, the error in measurement may cause biases and cast doubts on the validity of the results. Satisfaction and some of the hypothesized determinants are attitudes; hence, their measurement is prone to error. Consequently, the strength of the evidence in the previously mentioned studies could be affected by this methodological limitation.

Nonetheless, perceived measures in general, and quality in particular, appear most promising as determinants of satisfaction with public services. Further investigation of this issue is warranted. However, no specific methodological approach to examine the perceived quality of service was found within the public policy discipline.

According to Kettinger and Lee (1994), Marketing is a discipline within which “there has been considerable research on service delivery and customer satisfaction” (p. 742). It thus appears important to turn to this field of study to elicit a suitable course of action. In the following chapters, issues concerning the quality of service and its measurement will be examined. The last chapter of the literature review will deal with the measurement of satisfaction.

Perceived Quality of service

In the current sub-section, different issues surrounding the quality of service will be examined. The three steps followed are (a) establishing the distinctive attributes of services,

(b) finding the similarities between commercial and public services, and (c) modelling the perceived quality of service.

The concept of quality has been the focus of much attention in the Marketing literature. Crosby (1979) defined “quality as conformance to requirements” (p. 17). Juran (1974) preferred to envision quality as “fitness for use” (p. 2-2). These are only two of many possible definitions of the word *quality*. Indeed, “the term has been used in a bewildering variety of ways” (CEPPP, 1992, p. 3); hence, different persons may have individual interpretations of the concept of quality (e.g., Reeves & Bednar, 1994).

From a manufacturing perspective, quality is seen as the conformity to a pre-defined specification, which comprises sets of technical and functional norms to be achieved. Quality is then measured in terms of “conformance to requirements” (Zeithaml, 1988, p. 4). The absence of quality is measured through indicators, such as the frequency of the incidence of *internal* failures—flaws that are discovered during manufacturing—and *external* failures—faults that have surfaced after the customer took possession (Garvin, 1983). The traditional quality control approach is defensive and focuses primarily on preempting failures and eradicating defects (Garvin, 1987). In view of the negative connotation carried by the word *defect*, it has lately been replaced in the literature by *nonconformity* (Evans & Lindsay, 1996), a euphemism.

A good is a “physical object that has value to a buyer” (Brown & Fern, 1981, p. 205). Consequently, the above mentioned engineering approach to quality is particularly well suited for goods, because their physical attributes make them prone to discrete nonconformity. In summary, when assessing the quality of a good, there are several physical, intrinsic cues available for assessment, such as colour (Zeithaml, 1988).

Thus, there is an obvious “lack of isomorphism between goods and service quality” (Olivier, 1993, p. 69), because services possess distinctive attributes that set them aside from goods and renders inoperative the mechanistic definition of quality (Parasuraman, Zeithaml, & Berry, 1985). According to Chen, Gupta, and Rom (1994), “services differ

from goods in terms of how they are produced, consumed and evaluated” (p. 23). Indeed, a service is an activity, a social act, implying a direct contact between the provider and the recipient (Evans & Lindsay, 1996). Contrary to goods, performance constitutes the core offering of a service event (Moore, 1994). Services are processes—dynamic phenomena—that reveal difficult to grasp in view of the elusive nature of their underlying concepts (Grönroos, 1993). For services, a more restricted set of cues is generally available to assess quality (Richard & Allaway, 1993). When available, the only physical cues for a service relate to the provider in general, such as the premises and the equipment (Parasuraman et al., 1985), and are not sufficient by themselves to allow a comprehensive assessment of the services’ quality.

Consequently, quality for service is more difficult to measure, let alone define, than for goods (Johnston, Silvestro, Fitzgerald, & Voss, 1990). Also, quality judgements for services need to be formed differently than they are for goods.

Distinctive Attributes of Services

Differences between goods and services can be examined at the macro or at the micro levels. From the macro perspective, the service industry is characterized by specific traits. These traits are (a) a prevailing labour intensive profile, compared to the capital intensive nature of the goods industry; (b) a higher level of the interaction between the provider and the client; (c) the typical provision by multiple outlets (Berry, Parasuraman, & Zeithaml, 1988), and (d) compressed distribution channels (Olivier, 1993). More importantly, at the micro level these traits are (a) intangibility, (b) heterogeneity of output, (c) inseparability of production and consumption, (d) perishability, and (e) ownership.

Intangibility

Intangibles constitute a sizeable component of all services. Some services are *pure*, in that they are almost entirely intangible, while others contain a larger physical component. Because “they are performances rather than objects, precise manufacturing specification concerning uniform quality can rarely be set” (Parasuraman et al., 1985, p. 42).

Heterogeneity

In view of the prevailing labour content, each individual occurrence of a service activity develops in a slightly different manner, hence heterogeneous delivery, with an unpredictable effect on quality. In spite of a provider's efforts to maintain consistency through standardization, his inability to perform strict controls will result in observable variability in performance between employees, between customers, and between days. Consequently, "what the firm intends to deliver may be entirely different from what the consumer receives" (Parasuraman et al., 1985, p. 42). However, it has been suggested that excessive quality control would defeat its own purpose, interfering instead with appropriate delivery (Olivier, 1993).

In addition to the inability to control delivery, variability may occur as a result of the availability of several equally valid service avenues. Many services are delivered by professionals whose duties are to exercise considerable discretion in selecting the appropriate course of action according to their personal assessment of the needs and the conditions (Pollitt & Bouckaert, 1995). Consequently, similar customer functional requirements may be fulfilled with an array of appropriate solutions. The prescription of painkillers would be an example of this.

Heterogeneity may also reflect the variation in individual customer needs and expectations (Pettigrew, 1993). Sometimes, customers exercise a significant influence on the service act by setting some of the parameters of the delivery, such as specifications for a haircut (Parasuraman et al., 1985). This specific form of interaction between the provider and the recipient is generally referred to as *customization* and defined as "the tailoring of the service to the specific needs of the customer" (Harvey, Lefebvre, & Lefebvre, 1993, p. 486). The participation of one customer may also affect the service to others, such as the behaviour of partisans during a sports event.

When the customer participation exceeds expressing needs and implies some cooperation, like following a therapy, the "input-focused customer contribution" to the

service process has been coined as *co-production* (Lengnick-Hall, 1996, p. 797). This required cooperation between the provider and the recipient might impose a serious limit to the *designed-in* quality, which is more easily achieved for goods (Pollitt & Bouckaert, 1995). Olivier (1993) also asserted that the interpersonal dynamic that takes place during the service act enables the provider “to influence the consumer via expectancy shifts . . . as the service encounter unfolds” (p. 66). The same author posited that services demonstrate “dynamic properties” (Olivier, 1993, p. 66) that are not displayed by goods.

Inseparability

According to Kim (1990), inseparability of production and use is typical of most services, especially for labour intensive services. Production and consumption cannot be partitioned in distinct phases as is the case for goods, which are manufactured by the producer in a controlled environment and consumed at some later time (Parasuraman et al., 1985), with quality designed in unalterably (Pollitt & Bouckaert, 1995). In addition, quality control testing prior to sale is not feasible for services because they cannot be inventoried for future delivery (Parasuraman et al., 1985).

Perishability

Services are perishable because they cannot be stored. The service experience vanishes when the contact is removed. The empty seats on a plane flight cannot be inventoried for future use (Pettigrew, 1993). This inability to store this excess capacity when demand is weak, for use when demand has risen, constitutes a serious managerial problem (Pettigrew, 1993). In addition, the difficult “synchronization of supply and demand” induces a problem of reduced quality, because overwhelmed employees are much more “prone to mistakes” during periods of excess demand (Richard & Allaway, 1993, p. 61).

Ownership

For goods, purchase generally involves the physical transfer of the object. For services, the transaction only entails “the right to use, to access or to hire” (Buckley, Pass,

& Prescott, 1992), and only for a limited period. In summary, it can be concluded that services are unlike goods, hinting that quality of service must be assessed differently.

Similarities Between Commercial and Public Services

The discussion in the previous section was rooted in the Marketing literature, where the concept of service is implied to be generic. However, for the purpose of the current study, it is important to assess whether or not public services in general, and local services in particular, do form a subset of services. Otherwise, findings emanating from the commercial sector would have little bearing on municipal services.

Like commercial services, local services possess the basic attributes of services, namely intangibility, heterogeneity, and so forth, with varying levels of intensity. Supplying ecological compost containers and organizing economic missions are at the opposing extremes of the intangibility spectrum. Tax assessment may be done homogeneously, perhaps motivated by equity. However, library, parks and recreation, and tourism services are characterized by heterogeneous delivery (Das et al., 1995). Lineberry (1977) affirmed that co-production was observable for some local services. He mentioned fire protection, which is a “simultaneous function of the quality of the municipal fire department, the carelessness of residents, the private fire insurance industry, and the condition of local buildings” (p. 270). Community policing, which involves residents in the selection and the administration of the protection services, would constitute another example of this phenomenon. Inseparability is not absent from public services, as “government services are labor-intensive and tend to be produced and consumed simultaneously” (Battle & Nayak, 1994, p. 19). Perishability is also prevalent, as illustrated by police protection that exists only while the officers are present and dissipates when they leave. Finally, sitting on a park bench does not transfer title of the physical object. Thus, it may be concluded that municipal services are similar to commercial services on the basis of their basic attributes.

However, could municipal services differ from commercial services on the basis of some other attributes? Pollitt and Bouckaert (1995) argued that the three differences between commercial and public services are (a) the nature of the feedback link between provider and customer, (b) the existence of control services, and (c) the self-interest of public professionals. Donnelly, Wisniewski, Dalrymple, and Curry (1995) proposed (d) customer identification, as an additional distinction. Moreover, a fifth potential difference will also be examined in the current study, (e) instrumentation issues.

Feedback Link

Pollitt and Bouckaert (1995) posited that the link between provider and recipient is very direct for commercial services, with financial success used as a valid gauge of the quality of the service. They contrasted this with public services, where the economic link is not as clear, due to the free provision and the prevalence of rationing as a means to adjust demand with offer. Pollitt and Bouckaert (1995) added that the monopolistic position of the provider restricts the development of high expectations for services in view of the absence of reference. The authors specified that the monopolistic position precludes citizens from deserting and selecting another provider. The economic choice available to commercial consumers in view of poor service has been coined as the *exit option*, and is generally not available to citizens (CEPPP, 1992), in particular for publicly funded social assistance services (Gorey, Chandler, & Osmun, 1996).

While the above argument might hold for some services provided by higher jurisdictions, it could be argued that local governments are in no way monopolies or oligopolies. In fact, citizens can select which town they will live in. This decision process might entail a variety of factors, including the perceived quality of service prevailing in the prospective jurisdiction. When established in a town, dissatisfied residents can voice non-monetary feedback through complaints, pressures groups, and politics. In extreme cases, drawing on financial markets' efficiency theory, they can exercise *arbitrage* and change town.

Although there are costs involved in acting upon dissatisfaction, such as moving to another town, which might practically limit its extent, Stipak (1974) posited that residents might indeed use residential selection process to adjust their expectation level. He added that, in the context of contracted municipal services, residents of some cities may evaluate less favourably some specific aspects of municipal services, but their satisfaction might be similar to residents of other cities because they prefer and expect lower levels of service outputs. Consequently, although individuals at the lower levels of income might be more restricted in their options (Pollitt & Bouckaert, 1995), residents are generally free to select their town, making local services comparable to commercial services on that basis.

Some public services are not delivered in the context of a “clear market transaction—that is, one with a single customer paying for the service and enjoying the advantage conveyed by the transaction” (Picherak, 1987, p. 247). Here again, it might be argued that this might be more prevalent for services provided with higher level governments, as many municipal services are provided on a pay-for-use basis, such as recreation and permits. Furthermore, for real-estate owners, paying the tax bill constitutes a quasi-transaction with the city which, in counterpart, agrees to provide a predetermined basket of services.

Vehorn (1980) also took exception to the proposition of the absence of a market for public services. Although he admitted that public and private services are different in the way they are rationed, he argued that this distinction bears little consequence. Vehorn (1980) affirmed that private goods are rationed through the supply and demand pressures in the market, while public officials control the delivery of public goods through budgeting. He added that citizens who are dissatisfied with the level of service form the clientele of private entrepreneurs who launch initiatives to tap on the unfulfilled needs, such as private security agencies that compete with public police protection. The mixed market of recreation services would also qualify to illustrate this phenomenon. Consequently, Vehorn (1980) concluded that the mere availability of substitutes clearly constitutes a supply response to an unfulfilled demand and demonstrates that a market exists.

In summary, commercial and municipal services cannot be labelled as dichotomous on the basis of the market structure in which they are delivered.

Control Services

According to Pollitt and Bouckaert (1995), a second difference concerns *social ordering* services—the public services that are supplied as “part of the control apparatus of the state” (p. 13)—such as police and building inspection, sometimes against the will of the recipient. Such services may find few equivalents in the commercial services, but some counterparts can be suggested. For instance, the control mechanisms for obscene phone calls implemented by the telephone companies are designed to balance, and favour, the good of the general population against the desire of the actual user of the phone service, the obscene caller.

In addition, because quality measurement for services aims at assessing the experience during the service episode, it transcends the reasons that brought about the service. In most commercial services and many municipal services, the user seeks the service. In some commercial and local services, the service is delivered upon the recipient without request or sometimes involuntarily. Nonetheless, quality stems from the delivery. When examining quality in the context of “unwilling users”, the attention should be devoted to examining whether or not their “legitimate requirements” have been met (CEPPP, 1992, p. 6). Thus, a police intervention for speeding can be performed without quality if it is conducted with arrogance or abusive power. Alternatively, it can be conducted with quality if it is done expeditiously and with courtesy.

In summary, while there might be considerable difference between commercial and local services on what prompts the service experience, this difference has little bearing on the determinants of quality of the service experience.

Self-Interest of Professionals

Pollitt and Bouckaert (1995) posited that the third possible difference between commercial and public services concerns the propensity for public service professionals to

align their service delivery more with their vision of what constitutes social needs, as derived from professional beliefs, than the individual's real desires. An example would be vaccination (Pollitt & Bouckaert, 1995). This professional domination can be attributed to the fact that professionals "are largely self-regulated [and they] all have exceptional powers to define their patients' situation and their interests" (Bovaird, 1996, p. 40). This point parallels the previous one, except that self-interest fuels the underlying drive instead of social order.

Public "professionalized services" may share this "traditional culture" of provider "knows best" (CEPPP, 1992, p. 14). However that would be true of *all* professionals, whether they provide public or private services. Little differences can be found on that respect between lawyers and doctors. Furthermore, commercial services are *primarily* driven by the self-interest of the providers. This self-interest is generally referred to as profit. As a matter of fact, self-interest may be even more prevalent in commercial than in local services. At any rate, regulating agencies, such as the Bar Association, whose mandate is to protect the public, perform a check and balance between the professional's self-interest and the social need. The Better Business Bureau and consumer defence groups are commercial equivalents of such agencies.

In summary, it can be concluded that there is no difference between commercial and local services on the basis of self-interest of professionals either.

Customer's Identification

Donnelly et al. (1995) posited that for some public services there is a defined set of recipients who pay for the services they receive--albeit indirectly, similarly to commercial services. But, there are also recipients who pay nothing and payers who receive nothing, a situation which, they claimed, is not paralleled in commercial services.

However, many commercial services are beneficial not only to the payees, but also to others. An example of this is horticultural services performed on a lawn or flower beds. The landlord, who pays the gardener's fee, may be the prime recipient of the benefit.

However, the neighbours and every passer-by also freely benefit from the aesthete gains. Concerning those who pay without receiving any benefit, they are the exception rather than the norm in municipal services. Many services are provided on a need-only basis, such as the fire prevention services. A commercial parallel would be the casualty insurance services, which are paid for with the expectation that no claim will be submitted. So, the need and provision of such services is subjected to some probabilistic phenomenon and a tangible benefit exists even if it lacks salience. In summary, commercial and local services are also not conceptually different with respect to customer identification.

Instrumentation

Because the suspected differences between public and commercial services are not as prevalent as originally believed, they should have no bearing on the measurement of quality of service. Consequently, quality models and instruments emanating from commercial services may be considered valid for use with public services in general and local services in particular. Séguin (1991) asserted that all private sector quality indicators are fit for use with public services, thus encompassing local services as well, and underwriting this conclusion.

In addition to this conceptual demonstration, the empirical practice of researchers has been to use similar instruments for both types of services. Murdock and Schriener (1979) used a single yardstick for a comparative assessment of a mix of private and public services. More recently, Poister and Henry (1994) successfully used a common instrument to capture citizens' assessment of the perceived overall rating of quality for an array of public (including municipal) and private services.

Modelling the Perceived Quality of Service

Quality with services became a topic of great interest in the services literature around 1980 (Grönroos, 1993). Since then, several approaches have been used to attempt to explain the construct of quality as applied to services. Some of these approaches contain sizeable similarities that are reviewed below.

Grönroos (1983) proposed the Perceived Service Quality Model that inferred that the quality perceived by the consumer resulted from the comparison of expectations to experience with the service. High perceived quality occurs when experiences surpass expectations, while low perceived quality happens when the gap is negative. The quality construct is formed of two dimensions, *technical* quality and *functional* quality. Technical quality assesses the outcome—what the customer gets—while functional quality considers instead the process—how the customer gets it. This model is seminal in the service quality literature and can be credited with the bifocal paradigm that still supports this concept. First, it introduced the notion of confirmation/disconfirmation—expectations compared to perceptions. Second, it distinguished quality for services from that of goods by introducing delivery in addition to physical considerations.

Conceptualized in 1985, developed in 1988, and refined in 1991, the SERVQUAL scale was authored by Parasuraman, Zeithaml and Berry (PZB)³. These authors defined quality as the gap between perceived and expected service, measured within the scale as difference scores between two batteries of items. They posited that quality comprises five dimensions. These dimensions are (a) *Tangibles*, the physical cues of the service or facilities; (b) *Reliability*, consistency of performance; (c) *Responsiveness*, the readiness to provide the service; (d) *Assurance*, the courtesy of the agents and the confidence they inspire; and (e) *Empathy*, the attention provided to recipients (Parasuraman et al., 1988). This scale echoes Grönroos' (1983) service quality paradigm. The conceptual and empirical foundations of the SERVQUAL model and scale will be examined in subsequent sections.

Lethinen and Lethinen (1982) proposed a three-dimension model. These dimensions are (a) *physical* quality, (b) *interactive* quality, and (c) *corporate* quality—image. Physical quality is close to SERVQUAL's Tangibles dimension, while interactive quality can be associated to the Assurance and Empathy dimensions (Boyt, 1994).

³ A. Parasuraman, Valarie A. Zeithaml, and Leonard L. Berry can be credited with several joint publications on quality of service, many of which are referenced in this study. For simplicity, the acronym PZB will be used even when the authors appear in a different order, such as in Zeithaml, Parasuraman, and Berry (1990).

According to Swartz and Brown (1989), the corporate quality dimension is invariant in short to moderate time frames, is less situational, and less manageable than the other dimensions. So, corporate quality would not be assessed when services are examined. Furthermore, according to Boyt (1994), corporate quality is fully reflected through the five SERVQUAL dimensions. This model presents no significant advantage over SERVQUAL.

Garvin (1987) proposed eight components to quality, applicable to both goods and services. These dimensions are (a) *performance*, the primary operating characteristics; (b) *features*, the supplements to the basic functioning; (c) *reliability*, the absence of malfunctioning; (d) *conformance*, meeting standards; (e) *durability*, the value obtained while the product is operable; (f) *serviceability*, the possibility to obtain repairs and how they are performed; (g) *aesthetics*, a subjective appreciation of the physical cues; and (h) *perceived quality*, which is “inferred from various tangible or intangible aspects of the [service]”, such as reputation (p. 105). Garvin’s (1987) model is distinct in that it caters indiscriminately to goods or services. However, it does not place much emphasis on delivery, a key aspect of services.

Building on PZB’s ten determinants of service quality, Johnston, Silvestro, Fitzgerald, and Voss (1990) proposed a twelve-factor model to assess the quality of service. These factors are (a) *reliability*, of personnel and processes; (b) *responsiveness*, mainly response time; (c) *aesthetics*, the appearance of the facilities and agents; (d) *cleanliness/tidiness*, also focussed on the environment; (e) *comfort*, still environment relative; (f) *friendliness*, of agents; (g) *communication*, between agents and recipients; (h) *courtesy*, of staff; (i) *competence*, also of staff; (j) *access*, ease to attain facilities; (k) *availability*, of staff and service; and (l) *security*, which cover issues such as confidentiality. The work of these authors is interesting in that they utilized a bottoms-up approach to aggregate and classify quality measures used by ten companies. However, most measures are internal. In addition, no published study making use of this scale was found.

Finally, Brown (1990) proposed a bidimensional model. The two dimensions in Brown's (1990) model are *outcome* quality and *interaction* quality. According to their respective definitions, Brown's (1990) outcome quality is equivalent to Grönroos' (1983) technical quality and the former's interaction quality is also equivalent to the latter's functional quality. There is no significant difference between the two models (Boyt, 1994).

Each of these models puts into light a different subset of determinants of quality, with substantial overlap between the models. Sometimes identical, the models are at least very similar. This conceptual proximity may cause difficulty in finding criteria to discriminate between them and express a preference. In addition, all the above models show some potential for generalized use, at least at the conceptual level.

However, SERVQUAL presents significant advantages over the other models. First, it recognizes the specifics of services, as compared to goods, and it provides an external assessment focussed on the customer. Second, it fully encompasses Grönroos' (1983) and Brown's (1990) models and it shares many similarities with the other models. Third, although the models developed after SERVQUAL may be viewed as more specific and precise, SERVQUAL's five-dimension pattern appeals by its simplicity and parsimony. Fourth, in contrast to the other models that remain conceptual, SERVQUAL has been set into operation by means of a scale, with some known psychometric properties.

Consequently, the other approaches have received less interest in the literature. Indeed, the SERVQUAL scale has become the most widely known scale for the measurement of quality with services (Grönroos, 1993) and its most popular measure both in the literature and industry (Brown, Churchill, & Peter, 1993). Furthermore, "the SERVQUAL approach remains the *only*⁴ measure of service quality to date" (Olivier, 1993, p. 70). So much so that some studies that did not use the SERVQUAL scale have nonetheless used SERVQUAL's dimensions as a basis for item development (Boyt, 1994).

⁴ Emphasis added.

In summary, the search through the Marketing literature has permitted to identify a scale enjoying notoriety and specifically designed to assess the perceived quality of service. The search also revealed that although it was originally designed for commercial services, SERVQUAL might have some potential for use in the local service context.

Development of the SERVQUAL Scale

Conceptual Underpinnings

In 1985, PZB proposed a conceptual model to explain quality with services. The model was derived from a qualitative study on the concept of quality, relevant to four service categories: (a) retail banking, (b) credit card, (c) securities brokerage, and (d) product repair and maintenance. This array of services was designed to foster generalization of the results to most services, because “it represents a cross-section of industries which vary along key dimensions used to categorize services” (Parasuraman et al., 1985, p. 43). PZB assessed that their sample of services provided complete coverage because (a) retail banking and securities brokerage possess a *higher level of contact* with customers than the other two services selected, (b) product repair and maintenance is more *tangible* than the others, and (c) for credit card and product repair and maintenance, the service is delivered in discrete episodes, as opposed the *kind of continuous experience* that prevails with retail banking and securities brokerage (Parasuraman et al., 1985).

The data was gathered through “focus group interviews with [12 groups of] consumers and in-depth interviews with executives” (Parasuraman et al., 1985, p. 43). Three groups of recent users focused on each of the service categories. Participation was determined through quota sampling, predominantly in metropolitan areas of the U.S.A., controlling for age and gender. The in-depth interviews were conducted with a total of 14 senior executives from four industry leaders, one from each service category. The executives were selected from various areas within the firms, to represent a broad spectrum

of the different business components that may influence the quality of service in their respective organizations (Parasuraman et al., 1985).

The most important finding of the study was that similar and consistent results were obtained across the different samples. There were strong commonalities between industries in the executive interviews. Furthermore, “the focus groups revealed that, regardless of the type of service, consumers used basically similar criteria in evaluating service quality” (Parasuraman et al., 1985, p. 46). The apparent external validity of the findings encouraged the development of a general model for service quality.

PZB proposed a comprehensive model, the Service Quality Model, that construes quality as a series of five gaps between *what is supposed to be* and *what exists*, with regards to tasks in the service delivery chain. Four gaps are internal to the provider and they condition the fifth gap that is formed strictly at the consumer level. Gap 1 is the discrepancy between the actual consumer expectations and the firm’s management perceptions of the consumers expectations, indicating “that service marketers may not always understand what consumers expect in a service” (Parasuraman et al., 1985, p. 44). Gap 2 represents the misalignment between the firm’s management perceptions of the consumers’ expectations and the specifications that dictate the norms applicable to the service. This spread may result from constraints arising from resources or the market, but is also likely to stem from a lack of managerial commitment (Parasuraman et al., 1985). Gap 3 is the difference between the service specifications and the actual service delivery, which origins from “variability in employee performance” (Parasuraman et al., 1985, p. 45). Gap 4 results from the asymmetry between externally available information and the conditions of the actual service. This difference can result from unfulfilled (and sometimes fraudulent) provider promises. Or, it may result from a lack of consumers’ interest for features of the service that are predominantly in their best interest, such as safeguards (Parasuraman et al., 1985). Gap 5 is the discrepancy between expected service and perceived service levels. The Service Quality Model is illustrated in Figure 2.3.

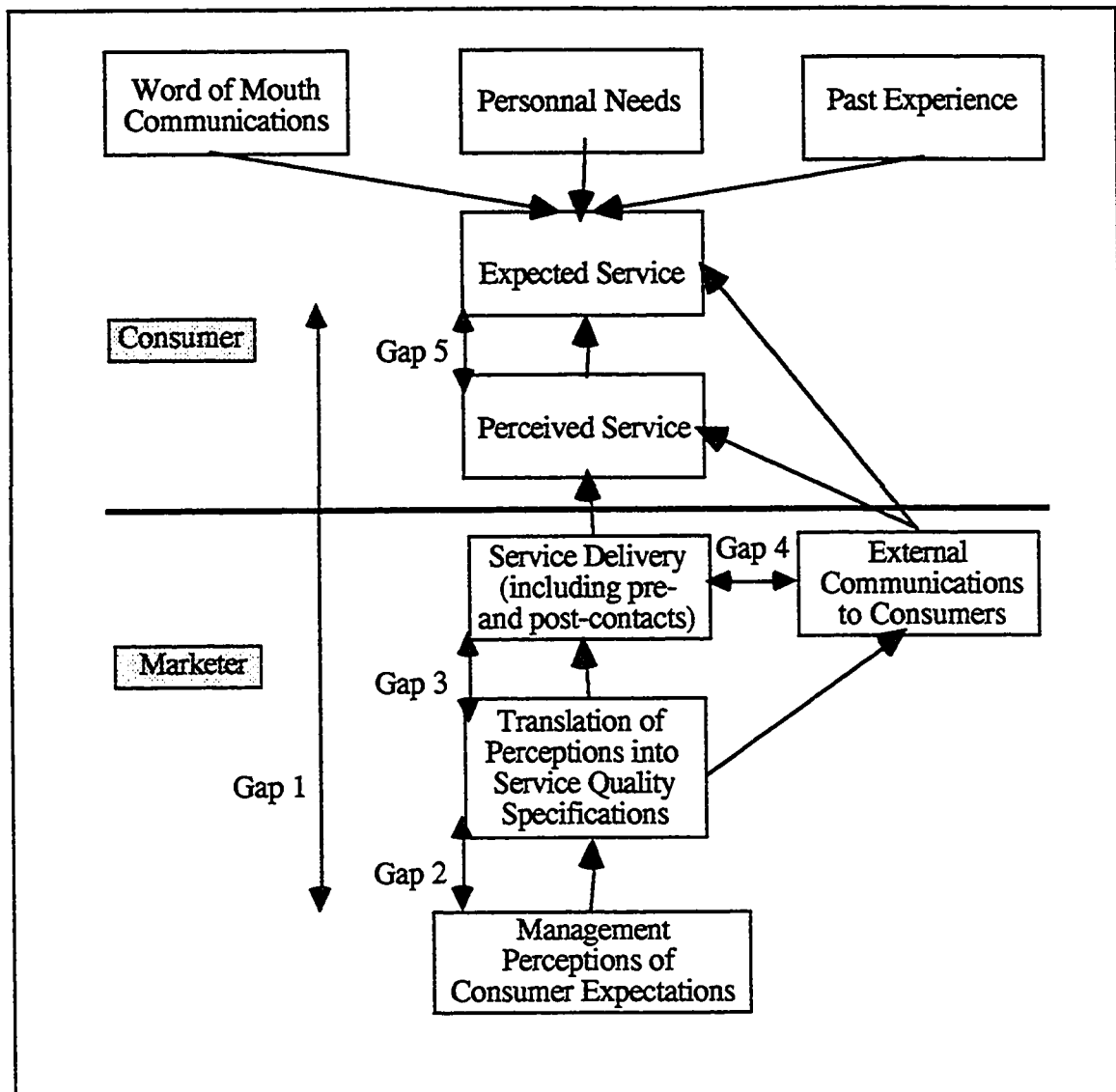


Figure 2.3. Service quality model⁵.

⁵ Parasuraman et al. (1985, p. 44).

The quality experience, as seen from the customer's perspective, is thus limited to Gap 5, which compounds the effects of the other four gaps. It is therefore a function of the size of each of the gaps and their directions. Gaps can also be positive. For instance, a positive result at Gap 5 is interpreted as delivery exceeding the customer's expectations (Parasuraman et al., 1985). Expectations are formed prior to the service experience and serve as a standard against which the performance will be compared (Boulding, Kalra, Staelin, & Zeithaml, 1993). The contrast, or disconfirming, role played by expectations can be tied to the fact that holding perceptions constant, higher expectations result in a lower level of gap quality (Boulding et al., 1993).

Quality occurs at a higher level abstraction than the intrinsic and extrinsic cues upon which it is formed, such that it can be generalized across brands and categories to allow comparisons (Zeithaml, 1988). In keeping, PZB's research revealed that "regardless of the type of service, consumers used basically similar criteria in evaluating service quality" (Parasuraman et al., 1985, p. 46). These criteria were labelled *determinants* of service quality. The ten determinants that they identified are listed, defined, and exemplified in Table 2.2. Overlaps, redundancy, lack of parsimony, as well as unequal relative importance were suspected; evaluation of these issues was deferred to future empirical research (Parasuraman et al., 1985).

In essence, to evaluate the quality of a service, a consumer compares what was obtained to what was expected, in a manner similar to evaluating the quality of goods. The only difference is that instead of forming the judgement predominantly on physical cues, the customer uses the previously mentioned determinants (Parasuraman et al., 1985).

The information acquisition process for a service (or a good, for that matter) can be classified according to the properties of this information. Nelson (1970) proposed a two-level framework comprising *search* and *experience* properties. Search properties refer to attributes that can be determined *ex ante* to the service experience and may therefore be used in the selection process. They generally refer to physical qualities, such as smell or

appearance. In contrast, experience properties can only be determined during or after the service act, such as the taste of the food.

Expanding this two-way classification system, Darby and Karni (1973) added a third property to the framework, *credence*. Credence properties refer to characteristics that are elusive to determination with certainty from the service experience, even *ex post*. An example would be the performance of an appendectomy, for which the patient “will have no different experience after the operation whether or not [it was required]” (Darby & Karni, 1973, p. 69). More examples could be derived from the health care and automotive repair contexts, where the technical nature of the service requires, for proper assessment, skills and knowledge that exceed those of the average customer. Darby and Karni (1973) also mentioned that the distinction between experience and credence might be blurred if the information becomes available through use, but only after considerable time has elapsed.

PZB’s determinants of service quality can then be aggregated into categories, along the three intrinsic information properties. Tangibles and credibility fall under the search caption. Access, communication, courtesy, reliability, responsiveness, and understanding/knowing the customer can be branded as experience. Finally, credence encompasses competence as well as security (Parasuraman et al., 1985). The determinants are classified along their respective properties in Table 2.2.

Service offerings are easiest to evaluate if they possess a high content of search properties, harder to evaluate if the experience properties prevail, and evaluation becomes close to impossible if credence properties are predominant. Unfortunately, services typically demonstrate few search properties, but high levels of experience and credence properties (Boulding et al., 1993). The absence of search properties and the lack of ability to assess the credence properties thus constraint the consumer to predominantly use experience properties in the assessment of services (Parasuraman et al., 1985).

Table 2.2.

Determinants of Service Quality

Information Property	Determinant	Definition (example)
Search	Credibility	Trustworthiness, believability, honesty; having the customer's best interest at heart (e.g.: firm's reputation).
Search	Tangibles	Physical evidence of the service (e.g.: physical facilities).
Experience	Access	Approachability and ease of contact (e.g.: convenient hours of operation and location of service facility).
Experience	Communication	Keeping customers informed in a language they can understand and listening to them; adjusting language for different customers (e.g.: explaining the service itself and how much it will cost).
Experience	Courtesy	Politeness, respect, consideration and friendliness of contact personnel (e.g.: clean and neat appearance of public contact personnel).
Experience	Reliability	Consistency of performance and dependability; performing the service right the first time; honoring its promises (e.g.: accuracy in billing).
Experience	Responsiveness	Willingness or readiness of employees to provide service; timeliness of service (e.g.: setting up appointments quickly).
Experience	Understanding/ knowing the customer	Making the effort to understand the customer's needs (e.g.: recognizing the regular customer).
Credence	Competence	Possession of the required skills and knowledge to perform the service (e.g.: competence and ability of the contact personnel).
Credence	Security	Freedom from danger, risk, or doubt (e.g.: confidentiality).

Note. From Parasuraman et al. (1985, p. 47).

Finally, PZB posited that perceived quality, as measured by Gap 5, existed along a continuum ranging from *totally unacceptable* to *ideal*, with satisfactory quality at some intermediate point. The values, scores when computed with the use of a scale, represent the direction and the magnitude of the discrepancy between expected service and perceived service (Parasuraman et al., 1985). In this context, expectations reflect the desires of the consumer, what the “provider should offer” (Parasuraman et al., 1988, p. 17), as compared to the consumer’s perception of the actual service.

Introduction of the SERVQUAL (1988)⁶ Scale

After developing the model, PZB presented an empirically tested instrument to measure perceived quality, the SERVQUAL (1988) scale. The development of the scale was initiated with the elaboration of 97 items covering the ten dimensions of quality. Each item was recast into a pair of statements. One statement was designed to measure the level of service desired from firms providing this type of service, while the other captured the perceived level of service quality provided by the particular firm under investigation. Statements about desire were gathered into an expectation battery and the others into a perception battery. The expectations battery was presented first in the instrument. Statements were positively and negatively worded, in similar proportions, following the prevailing procedures for marketing scales development (Parasuraman et al., 1988). The measures consisted of seven-level, Likert-type scales, anchored at extreme values (Parasuraman et al., 1988).

The self-administered two-part instrument was first tested on a quota sample of 200 respondents, 40 customers from each of the four categories used in 1985, plus long-distance telephone. All the responses were pooled. Difference scores were computed for the 97 pairs of statements, where perceived quality (Q) is the difference between

⁶ The scale was designated as SERVQUAL by the authors. However, PZB modified the scale somewhat in 1991, notably by making all statements unidirectional. To avoid confusion between the 1988 and the 1991 versions, SERVQUAL (1988) will refer to the earlier version and SERVQUAL will point to the revised scale. For matters in which the scale is invariant between versions, such as the number of factors and the definitions of the constructs, SERVQUAL will be used.

perceptions (P) and expectations (E), such that $Q = P - E$. The factors were extracted by principal axis factoring (PAF) with an OBLIMIN rotation. The coefficients α (Cronbach, 1951) for the ten dimensions ranged from .55 to .78.

Several waves of item selection analysis resulted in a seven-factor, 34-item instrument, with α values for factors ranging from .72 to .85. In this purified version, only five of the original dimensions formed distinct factors: Tangibles, Reliability, Responsiveness, Understanding/Knowing the customer and Access. The other five dimensions collapsed into two factors, labelled D4 and D5. The total-scale reliability was .94.

PZB then initiated a second stage of scale development. Customers of four large and well-known service firms were surveyed in person. The service sectors represented by these firms were banks, credit cards, appliance repair and maintenance, and long-distance telephone. Between 175 and 187 recent users for each concern, selected by quota in a shopping mall, were subjected to the reduced 34-item self-administered instrument (Berry et al., 1988). The questionnaire also included a global quality assessment item, an item to assess the propensity to recommend the firm and an item covering complaints about poor service.

Scale purification resulted in a further reduced 22-item per battery instrument, the SERVQUAL (1988) scale (Parasuraman et al., 1988). The specifications of the scale are reported in Table 2.3. The new scale comprised only five factors: D4 and D5 having collapsed into Assurance, while Understanding/Knowing the customer and Access were joined to form Empathy. Only nine of the 22 remaining items were worded negatively (Carman, 1990). The patterns observed were similar across the four samples. A second analysis of the data gathered during the first stage, using this new structure, yielded similar results, confirming the appropriateness of the five-factor simple structure (Parasuraman et al., 1988).

Table 2.3.
Specifications of the SERVQUAL (1988) Scale

Dimension	N of items	Definition
Tangibles	4	Appearance of physical facilities, equipment, personnel, and communication materials.
Reliability	5	Ability to perform the promised service dependably and accurately.
Responsiveness	4	Willingness to help customers and provide prompt service.
Assurance	4	Knowledge and courtesy of employees and their ability to convey trust and confidence.
Empathy	5	Caring, individualized attention the firm provides its customers.

Note. From Zeithaml et al. (1990, p. 26).

PZB assessed the psychometric properties of SERVQUAL (1988) by considering reliability and validity. The factors yielded Cronbach α values for the factors ranging from .52 to .87. Tangibles produced the lowest values (average = .60), while Assurance and Reliability yielded the highest values (averages = .84 and .79). *Face-or content-validity* was appreciated qualitatively. Given that the dimensions were explicated thoroughly and the scale items fully represented the constructs' domains, the scale was found to satisfy the conditions of content validity. In addition, investigation of the relationship between the scale and the conceptually related items, namely overall quality, propensity to recommend the firms, and reported problems arising from service, revealed that propensity to recommend and absence of reported problems were related to high-perceived quality. This further supported the validity of the scale (Parasuraman et al., 1988).

PZB also attempted to gauge the relative importance of the individual dimensions. To do so, they performed a regression of the individual dimension scores on the total-scale score. Dimension scores are computed by averaging the composing item scores; the total-scale score is the average of the 22 scores. The regression coefficients were rather small

and not all significant. This lack of relationship was attributed by PZB to multicollinearity between the measures.

In the second phase of the study, PZB included items eliciting the importance of the dimensions. Except for Tangibles, all dimensions had average scores in excess of 9 on a 10-point scale and Reliability was the most important dimension for the majority of respondents (Berry et al., 1988).

The authors suggested that the elicitation of a *universal* scale caused the elimination of items which use may be desirable when investigating a particular industry or firm. They also suggested that “items under each of the five dimensions can be suitably reworded and/or augmented to make them more germane to the context in which the instrument is to be used” (Parasuraman et al., 1988, p. 28). In conclusion, PZB inferred that the scale would provide good results when used to assess the quality of service from a wide range of service areas (Parasuraman et al., 1988).

Critiques of the SERVQUAL (1988) Scale

Several researchers investigated the properties of SERVQUAL (1988). A sample of these studies is reviewed below (cf. Asubonteng, McLeary, & Swan, 1996 for an exhaustive coverage). These studies and the SERVQUAL (1988) scale are put into perspective in Table 2.4.

From a conceptual perspective, Johnston et al. (1990) observed some ambiguity in the definitions of the dimensions. They claimed that it is not the Tangibles present in the service offering that are a quality determinant, but rather their appreciation in terms of the quality factor. In addition, they claimed that the other four dimensions also point to the tangible elements. For example, the accuracy of records is not related exclusively to data entry, but also receives some contribution from the physical support elements.

Babakus and Mangold (1989) used adapted scales to assess perceived quality in a health-care setting. An exploratory factor analysis (EFA) performed on each battery revealed that both were unidimensional. Babakus and Mangold (1989) also formed

composites with the items along the five dimensions in each battery and used these bundles as indicators to test a two-factor correlated model. Babakus and Mangold (1989) found that the factor loading estimates and the fit were good. However, using composite scores in a partially aggregated model constitutes an admission that the scales are multidimensional, with the composites reflecting the underlying individual constructs.

Carman (1990) provided another assessment of SERVQUAL (1988). Construct validity of the scales was not confirmed, although many items loaded on the same factors as in PZB's (1988) study. The emerging dimensions in the four settings study were five to seven factors simple patterns when retaining factors with eigenvalues ≥ 1 .

Finn and Lamb (1991) reported a bad fit of the five-factor model with data collected from customers of retail stores. They reported an adjusted goodness of fit index (AGFI) for the model of .85, barely short of the .90 popular threshold. In addition, the reliability measures of the factors were comparable to PZB's (Finn & Lamb, 1991).

Babakus and Boller (1992) examined the use of the SERVQUAL (1988) scale with residential customers of an electric and gas utility company. The dimensionality of the scale was examined through an *EFA*, using the PAF factor extraction algorithm and oblique rotation, constraining the solution to five factors. While the reliabilities of the five *a priori* factors are similar to PZB's, the expected pattern was not met. A confirmatory factor analysis (*CFA*) for a five-factor correlated model and another one for a second-order factor model with five first-order factors failed to generate adequate fit.

Refinement of the SERVQUAL Scale

PZB came out with a refined SERVQUAL scale (Parasuraman, Berry, & Zeithaml, 1991a; Zeithaml, Parasuraman, & Berry, 1990), which paid attention to the problems encountered with the first version of SERVQUAL and, to a lesser extent, countered their critiques.

Table 2.4.

Some Critiques of the SERVQUAL (1988) Scale

	PZB (1988)	Babakus & Mangold (1989)	Carman (1990)	Finn & Lamb (1991)	Babakus & Boller (1992)
Service areas	Bank, credit card, appliance repair, l.d. telephone	Hospital	Dental clinic, placement centre, store, hospital	Four retail stores	Utility
Sample size	4 x 200	443	74 to 600+	58 to 59	689
Sampling and reply rate	Quota	Stratified, 22%	n.r.	Quota, 32%	Systematic, 29%
Administration	Self admin., in mall	Mail	Self-admin, on site	Telephone	Mail
Questionnaire	22 items, two batteries	15 items, two batteries	16-40 items, 1&2 batteries	As PZB	As PZB
Response scales	7-point Likert	5-point Likert	n.r. ^a	5-point Likert	n.r. ^b
Negatively worded items used	Yes	No	Yes	Yes	Yes
Gap scores used	Yes	No	Yes (1/4)	Yes	Yes
Analysis and rotation	EFA, PAF, OBLIMIN	EFA, n.r. ^c , CFA	EFA, n.r. ^d , PROMAX	CFA	EFA, PAF, oblique, CFA
N of Factors	Five	One per battery, partial aggregation	Five to seven ^e	Poor fit with five factors	One
Reliability of factors	.52 to .87	.90 and .96	.51 to .94	.59 to .83	.67 to .83
Issues raised		Dimensions, negatively worded items, gap scores, universality	Dimensions, negatively worded items, gap scores, universality, expectations, importance of dimensions	Dimensions	Dimensions, negatively worded items, gap scores, universality

Note. n.r. = not reported.

^aSeven-point Likert according to Parasuraman et al. (1991a). ^bSeven-point Likert according to Parasuraman et al. (1991a). ^cAsubonteng et al. (1996) mentioned PAF and oblique rotation.

^dAsubonteng et al. (1996) mentioned PAF. ^eParasuraman et al. (1991a) mentioned six to eight dimensions.

For the purpose of pre-testing, the earlier instrument was administered to a sample of 300 customers from a participating firm, with a reply rate of 23%. The results prompted the refined version. This version differs from the previous one in that (a) the expectations referred to what the best firms do, not to what a firm should do; (b) all items had been positively worded; (c) two items were changed; and (d) the importance of the dimensions was measured distinctly.

PZB mentioned that the change in the wording of expectations aimed to correct the “unrealistically high expectation scores” (Parasuraman et al., 1991a, p. 422) obtained when eliciting the expectations as to what the firms *should* do. PZB found several problems associated with the use of negatively worded items. First, they showed more variability, which was attributed to confusion on the respondents’ part. Second, the reliability of such items was lower. Third, the management of participating firms was not entirely pleased with the negatively worded items, a statement that parallels that of Carman (1990). Consequently, all items were made unidirectional.

In the Tangibles section, an item dealing with the communications material was substituted to another item. In the Assurance section, an item that was actually referring to “behind-the-scenes support” (Parasuraman et al., 1991a, p. 423), a phenomenon that is transparent to most customers, was replaced by an item on the competence of employees. A few other cosmetic changes were made to some items.

The authors also included a constant-sum 100-point scale to directly measure the relative importance of the dimensions. This scale was introduced to allow the computation of weighted scores (Parasuraman et al., 1991a).

The empirical validation was performed by mail on random samples of 1800 to 1900 customers of a telephone company, two banks and two insurance companies. The survey yielded a response rate of 21%.

Constraining the solution to the five-factor *a priori* pattern, the prediction was not verified for difference scores. In fact, the five samples showed identical exceptions to it.

Tangibles broke up into two dimensions (items 1 & 2, items 3 & 4), Responsiveness and Assurance loaded on the same factor, while Empathy and Reliability behaved as predicted. Also, the inter-correlation between the factors was higher than in the first study. When the batteries were examined separately under the five-factor constrained solution, expectations revealed closer to a four-factor pattern: Tangibles, Responsiveness/Assurance, Reliability and Empathy. Perceptions had a factor pattern similar to difference scores, with Tangibles split as above. Otherwise, in both instances the item loading estimates on the factors were identical to that of the composite scores. The split of Tangibles was interpreted to mean that customers assessed the quality arising from physical cues differently “than they do the appearance of employees/communication materials” (Parasuraman et al., 1991a, p. 429).

PZB hypothesized that the collapse of Responsiveness and Assurance into a single factor was a method effect attributed to the constrained solution. They assumed that, because Tangibles take up two factors and Reliability as well as Empathy one each, there was only one factor left for Responsiveness and Assurance (Parasuraman et al., 1991a). An *EFA* was thus performed without constraining the solution on the perceptions battery and the difference scores. They both revealed six-factor patterns, with the now familiar Tangibles split. In both instances, three of the other factors received items loading estimates according to the expected pattern. For perceptions, the Responsiveness factor was present, with considerable overlap with Reliability and Assurance. It merged with Assurance when difference scores were used. However, when PZB examined the results of the fixed-sum scale, they discovered that a significant difference existed between scores relating to the importance of the Responsiveness and Assurance dimensions, which confirmed the distinction that respondents make between these seemingly confused constructs. They also found that Reliability was rated as the most important dimension (Parasuraman et al., 1991a).

From these various analyses, PZB concluded that while more overlapping between the dimensions was evidenced, the five-factor pattern was verified, except for the

partitioning of Tangibles (Parasuraman et al., 1991a). Reliabilities for this empirical validation were higher than for the first version. This was interpreted as evidence that the refinements had impacted favourably on the cohesiveness of the dimensions (Parasuraman et al., 1991a). Validity was assessed to be achieved.

PZB also mentioned that perceptions *alone* had a stronger relationship to the criteria than the full scale, questioning both the utility of expectations and the gap scores. However, they stressed the diagnostic value of having both batteries distinct, because trends in each can be monitored over time (Parasuraman et al., 1991a, p. 433).

Critiques of the SERVQUAL Scale

Since the introduction of the revised instrument, SERVQUAL has been used in numerous studies conducted in a wide range of different fields, many of these studies subjecting the scale to some form of assessment. In view of the difficulties associated with tracking down the use of a particular procedure in studies across disciplines, no effort has been made in the present study to census all of them. Hence, an *ad hoc* sample of studies that used SERVQUAL is reported *infra*. The studies mentioned in this second round of critiques and the SERVQUAL scale are put into perspective in Table 2.5.

Cronin and Taylor (1992) obtained random samples of 175 to 189 complete replies from recent customers of firms in the following industries: banking, pest control, dry cleaning, and fast food. The five-factor pattern of SERVQUAL was not evidenced in any of the four samples when tested through a CFA. Cronin and Taylor (1992) hypothesized that the perceptions battery alone, which they called SERVPERF, was a better instrument than SERVQUAL to capture the attitude towards quality.

Teas (1993) found many conceptual and operational difficulties with PZB's gap scores and the measurement of expectations. Diverging from PZB's approach on these issues, he proposed the Evaluated Performance model (EP), based on the absolute value differences between performance and the *classical ideal point* expectations—a finite, optimal expectation level beyond which performance will entail displeasure, which he measured

with a seven-point semantic differential scale. The EP model provided significantly better results than SERVQUAL.

Brown et al. (1993), using a convenience sample of 230 respondents, compared difference scores to a *direct* measure assessing perceptions and expectations with a single item. While it reinforced the evidence against difference scores, it turned out to be only a modest improvement. In addition, they asserted that the scale revealed unidimensional, but they did not specify their assessment criterion. Three factors could have been retained on the basis of latent roots in excess of one: the first factor accounted for 51% of the variance and the other two totalled an additional 14%.

Recent Developments

Pursuing their quest, PZB (1991b) conducted a subsequent qualitative research with 16 customer focus groups, involving business as well as *end-customers* (i.e., consumers), and focussing specifically on customer expectations. Adding specificity to their conceptual model, PZB hypothesized that Reliability is centred on the outcome of the service, and evaluated by the customer *ex post*, while the other four dimensions focus on the service process, and are assessed concurrently to delivery. PZB mentioned that “Reliability is the most important dimension in *meeting*⁷ customer expectations, [but] the process dimensions . . . are most important in *exceeding*⁷ expectations” (Parasuraman et al., 1991b, p. 41). Their findings also suggested that expectations are dual-levelled, dynamic, and could be best expressed as a *zone of tolerance*, ranging from *adequate* to *desired* levels. These zones of tolerance vary in the levels of their bounds. The desired service expectations for Reliability are higher in view of their importance and would rise with experience. To the contrary, the adequate level bound are lower. In addition, the adequate level would rise with previous poor experiences (Parasuraman et al., 1991b, p. 43). Hence, good experiences would raise the upper bound of expectations, while poor experiences would raise its lower bound.

⁷ Emphasis added.

Table 2.5.
Some Critiques of the SERVQUAL Scale

Criteria	Parasuraman (1991)	Cronin & Taylor (1992)	Teas (1993)	Brown et al. (1993)
Service areas	Telephone, bank, insurance	Bank, pest control, dry cleaning, fast food	Discount store	Financial institution
Sample size	5 x \pm 1850	175 to 189	120	230
Sampling and reply rate	Random, 21%	Random	Random	Convenience
Administration	Mail	At home interviews	Personal	Self-admin.
Questionnaire	Two batteries, importance scale	Similar to PZB	Several scales, two items per factor	Direct Measures
Response scales	7-point Likert, constant-sum	7-point Likert, semantic dif.	7-point Likert, semantic dif.	7-point Likert
Gap scores used	Yes	No	Yes	No
Analysis and rotation	<i>EFA</i> , oblique	<i>EFA</i> , n.r. ^a , OBLIMIN, <i>CFA</i>	Correlation	<i>EFA</i> ,
<i>N</i> of factors	Six	One	One	One (three)
Reliability of factors	.80 to .93			
Issues raised		Dimensions, gap scores, universality	Gap scores, expectations, importance of dimensions	Dimensions, gap scores

^aAsubonteng et al. (1996) mentioned PAF.

In a further qualitative study, PZB built on the desired and the adequate levels as components of expectations (Zeithaml, Berry, & Parasuraman, 1993). They replaced the single quality concept by two distinct measures of service quality, each grounded to a specific component for expectations. *Perceived service superiority* was defined as the gap between perceptions and the desired service level. The desired service level fitted with the ideal, normative approach to expectations, but blended what the firm can deliver to what the customer believes he should receive. It is described as the “level of performance the customer ought to receive, or deserves, given a perceived set of costs” (Zeithaml et al., 1993, p. 6). This index was close to the original scale. More different was the gap between perceptions and the adequate service level, which was defined as *perceived service adequacy*.

This more recent version of the scale was subjected to empirical validation (Parasuraman, Zeithaml, & Berry, 1994a). The items were generally identical to the 1991 version. Alternative indices, using direct and difference measurements, were constructed with perceptions and the two forms of expectations. PZB observed that the perception-only scale had the highest predictive power, with direct measures next and the difference scores lowest. Yet, they found no difference in the psychometric properties of the gap scores and the direct measures. Both instances showed similar results: the five-factor pattern was not verified, Tangibles and Reliability emerged, and there was considerable overlap for the other three dimensions.

Issues with the SERVQUAL Scale

In this section, the various issues that were raised in the previously mentioned studies about either version of the SERVQUAL scale are examined in more detail. These issues are (a) negatively worded items, (b) dimensionality, (c) universality, (d) relative importance of the dimensions, (e) expectations, and (f) difference scores.

Negatively Worded Items

The use of items worded in a negative fashion is a recommended practice (Alreck & Settle, 1995). Alternating directional statements are intended to prevent the prevalence of yea/nay-saying (Alreck & Settle, 1995) and halo effects (Carman, 1990).

Substantiating the exclusive use of positively worded items, Babakus and Mangold (1989) felt that “the mixture of negatively and positively worded statements created confusion and frustration on the part of respondents” (p. 195). Carman (1990) supported the proposition that a mixture of negatively and positively worded items may confuse respondents, added that confusion is likely to increase with the length of the instrument, and mentioned that sponsors would not condone research material that would suggest negative proposition about their business, a position supported by PZB (1991a). Babakus and Boller (1992) found that negative statements caused the batteries to split along the item wordings, but gap scores were free of this method effect.

In conclusion, while there is a debate about the conceptual advantages and disadvantages surrounding the use of bifocal statements, the empirical tests produced evidence mostly against the use of reversed items. The refined version of SERVQUAL used only single-directional items. The increased levels of reliability found in the factors was attributed to this change (Parasuraman et al., 1991a).

Dimensionality

Dimensionality is concerned with determining if a construct is formed of only one “or a cluster of related attitudes” (Vogt, 1993, p. 70). From an empirical perspective, the dimensionality of the scale hinges on determining the number of factors that are supported by the data. In this context, a scale is said to be *unidimensional*, if all items load strongly on a single factor, or *multidimensional*, if there is exclusive loading on one of many factors.

During scale development, PZB found that the five conceptual dimensions were empirically supported by an equal number of factors emanating from the data sets.

However, many replication studies failed to reach the same conclusion. Babakus and Mangold (1989) found each battery to be unidimensional, when difference scores were not computed. However, they computed composite scores along the five PZB's dimensions for purpose of further statistical analysis and found evidence in favour of a partially aggregated second-order factor model, pointing instead to a more complex factor pattern. Also not computing difference scores, Carman (1990) found five to seven factors in his various samples. When gap scores were used, the five-factor pattern was attained in some studies (Ford, Joseph, & Joseph, 1993; Parasuraman et al., 1988). In a *CFA* study, the fit index score of a five-factor model failed to attain the customary threshold, but was close (Finn & Lamb, 1991). Otherwise, a unidimensional pattern was found (e.g., Babakus & Boller, 1992).

To explain why the five-factor pattern was not always supported, the conditions surrounding the tests in each study must be considered. Sometimes, the scale was subjected to considerable modifications. Although PZB suggested that it was appropriate to customize the scale, "to fit the characteristics or specific research needs of particular organizations" (Parasuraman et al., 1988, p. 31), results from tests on heavily modified batteries may not evidence the behaviour of the unaltered scale.

Secondly, a wide range of statistical tools, including both *EFA* and *CFA*, were used to obtain the results. In some cases, *EFA* with the PCA extraction algorithm was used, although the usefulness of PCA in factor extraction has been questioned (Hubbard & Allen, 1987). These various techniques are founded on different methodological underpinnings and produce unbiased results only when method specific modelling and distribution assumptions are met. It is doubtful that data sets hypothesized to belong to the same general population simultaneously fulfil all these diverging conditions. Hence, some methods' assumptions may not have been met, making the assorted results dubious.

Third, concepts take precedence over the data in the elaboration and validation of theories, because measures are subject to sampling effects that are independent from the

theoretical underpinnings. According to Babakus, Pedrick, and Inhofe (1993), a single factor does not disprove a multidimensional construct. Babakus et al. (1993) added that the lack of verification of the five-factor pattern in a study would be data dependent, without necessarily invalidating the theory.

Consequently, the hypothesized factor pattern has received mixed evidence. However, there is no reason to disbelieve that it accurately explains the phenomenon as theorized and some to believe that it does.

Universality

SERVQUAL' universality as proposed by PZB (1988) was challenged with respect to the complexity of services. This questions its applicability across service functions, because the dimensionality of the scale may be a function of the type of service.

Carman (1990) hypothesized that the Tangibles dimension in one of his samples may have split along different episodes of the service provision. The more complex, multifaceted services would thus result in an increased set of factors.

Having found the scale to be unidimensional, Babakus and Boller (1992) suggested that the five dimensions of the service provided by a utility, although being present, lacked salience to the consumer. A follow-up study on a different sample of customers of the same utility lead to a similar conclusion (Babakus et al., 1993). Babakus and Boller (1992) suggested that utilities typically deliver a steady, low profile service to *contented* customers, who rate the firms' quality evenly across dimensions. This results in high correlation between the scores and fewer factors extracted. Consequently, the true factor pattern could not be distinguished and a unidimensional pattern surfaced.

So, for Carman (1990) the complexity of the service caused more dimensions to appear, while for Babakus and Boller (1992) the scale became unidimensional because the service was simple and straightforward. These two findings converge to the idea that the number of factors emerging from the data would be a function of the complexity and/or salience of the service to the consumer.

Building on Babakus et al.'s (1993) argument about simple services, it could also be hypothesized that the magnitude and the variability of the attitude are a function of the complexity/salience of the service. This would entail that for simple services, perceptions and expectations are close. The quality scores would be small in magnitude, with little dispersion, and would be similar for all firms. For complex services however, each facet's visibility raises the potential for divergent perceptions and expectations. The scores would then demonstrate a larger range, with more variability for all firms in the industry. The magnitude and the sign of the average scores would thus reflect the achievement of the individual firms.

To remedy this phenomenon, Carman (1990) proposed that multifaceted services should be measured in phases. In addition, PZB can be credited for suggesting to customize the scale to context specifics by removing or adding items, while keeping the scale's basic structure and paradigm. This avenue has led to the development of derivative or industry focused instruments, such as SERVPERF (Cronin & Taylor, 1992) and LODGQUAL (Thompson & Getty, 1994).

In summary, the empirical findings and their substantive interpretation raise some doubts as to the ability of SERVQUAL to provide equally meaningful information about quality in all industries.

Relative Importance of the Dimensions

Originally raised by Carman (1990), the issue of the relative importance of the dimensions comprises two aspects: How to measure the relative importance of the dimension? and, Is it worth it?

Concerning measurement, PZB (1988) attempted without success to derive the weights from the scale itself, through regression. Multicollinearity was targeted as the problem. Carman (1990) posited that the importance of a service attribute is different from its expected level and that it must be measured distinctly, using separate items to gather information on this topic. The use of a constant-sum scale for this purpose can first be

traced to Crompton and Mackay (1989), whose study will be reviewed in the next section. This practice was adopted by PZB and by other researchers thereafter. The issue was thus resolved in favour of a distinct measurement.

On the usefulness of the importance weights, while PZB advocated the universality of the scale, they always assumed a varying weights structure for the dimensions reflective of customer segmentation (Parasuraman et al., 1988), or perhaps of industry particulars. Carman (1990) had also hypothesized unequal weights, but his results from the dental clinic survey revealed homogeneous importance across the five dimensions. However, other studies have evidenced unequal weights (Crompton & Mackay, 1989; Cronin & Taylor, 1992; Parasuraman et al., 1991a; Zeithaml, Berry, & Parasuraman, 1996). In conclusion, the evidence is mixed with unequal importance leading. In addition, the absence of difference between dimensions' importance only represents a special case of the general weighing scheme. Thus, using a valid measure of the relative importance only adds value to the scale information.

In conclusion, a distinct measurement of the relative importance appears warranted in view of its enrichment of the information otherwise captured by the scale.

Expectations

The measurement of expectations raises four related, but distinct issues. The first issue concerns the utility of measuring expectations at all. The second one points to the instability of expectations as a function of experience. The third issue is related to the appropriateness of measuring expectations *ex post*. The last issue concerns the method used to consider expectations and perceptions simultaneously.

Several authors have gathered evidence that suggests that perceptions alone define quality, reducing the utility of introducing expectations at all into the equation (Babakus & Boller, 1992; Boulding et al., 1993; Carman, 1990; Cronin & Taylor, 1992). In fact, PZB admitted that, even in their own studies, perceptions revealed to be better than gap scores as determinants of criterion validators (Parasuraman, Berry, & Zeithaml, 1993).

The apparent redundancy of expectations in gap scores can be attributed to a large extent to a method effect. As will be discussed *infra*, the perceptions component exercises a stronger effect on an index measure than the other component, which induces the composite to behave similarly to its influential component. Teas (1993) produced a different index in which both components appeared to carry some weight, although the index itself may be questionable because it used absolute difference scores.

Furthermore, PZB insisted that the simultaneous presence of both components conferred to the scale a higher *diagnostic value* than a perception-only scale (Parasuraman, Zeithaml, & Berry, 1994b). However, it was noted that the interest in the possible richness and diagnostic value of SERVQUAL should be subordinated to confirmation of its psychometric properties (Cronin & Taylor, 1994).

Doubts concerning the usefulness of the expectations component in composite measures are not exclusive to the development of SERVQUAL. In a study conducted by Dorfman (1979) on satisfaction with outdoor recreational experiences, the difference scores between preferences and expectations for environmental or personal “factors . . . related to camping experiences” (p. 487) failed to correlate highly with satisfaction measures, while perceptions alone did. In summary, expectations may not be useless; they may be under-utilized.

The second issue concerns the instability of expectations as a function of experience. Indeed, the expectations of a new customer may differ from those of a customer who has used the service before (Waddell, 1995). Carman (1990) hypothesized that new customers would have unrealistically high expectations because they are not familiar with the service. He proposed to consider only the averaged expectations of experienced customers. In paradox, PZB’s qualitative study found that expectations *raised* with experience (Parasuraman et al., 1991b), a finding that concurs with the *ratchet effect* paradigm of upward-mobile, dynamic expectations. So, there appears to be a consensus that the expectations of experienced customers are more meaningful to the firm. However,

there is uncertainty as to the direction in which expectations are moving as a result of experience.

With regards to the third issue, simultaneous administration of both batteries assumes a static model, without interaction between expectations and perceptions during the service experience (Grönroos, 1993). More so, “it is reasonable to assume that the expectations of a given customer change continuously . . . during the service encounter” (Grönroos, 1993, p. 59). Having recognized this difficulty, Carman (1990) used a separate administration. However, dual administrations are cumbersome, so Carman (1990) recommended periodical gathering of this data, for multiple use. Grönroos (1993) disagreed with this method because the expectations of consumers measured *ex ante* may not be “the expectations to which they compare their experiences” (p. 56). He added that a dynamic model of the relationship is not yet developed and that measuring perceptions alone may be sufficient.

In an opposing view on this issue, Boulding et al. (1993) posited that the expectations considered in SERVQUAL point to *normative*, ideal expectations, fostering stability over time. Thus, the timing of the measurement of expectation becomes irrelevant, because the respondents’ expectations are fixed. This approach also meets the monotonicity assumption required *infra* by Stipak (1979a). Also concurring that expectations can be measured “in retrospect”, Olivier (1981, p. 39) mentioned that prior measurement are nonetheless preferable and that concurrent measurement may cause interaction between both measures. He also asserted that in certain circumstances posterior measurement is even preferable.

Thus, a concurrent administration of both SERVQUAL batteries would seem appropriate. Notwithstanding the issue of the method to link expectations to perceptions, the use of a distinct measure of expectations appears to add significant value to the information. The questions raised on how to examine both measures in a joint manner leads to the issue of difference scores, which is covered in the following sub-section.

Difference Scores

Some researchers find it important to establish congruence, or lack of congruence, between measures that are commensurate. Commensurate measures are “measures express[ing] both components in terms of the same content dimension” (Edwards, 1994, p. 72). One available method, difference scores or gap scores, is a bivariate measure formed by an index of the algebraic difference between the two commensurate component measures. In view of their “seductive face validity” (Johns, 1981, p. 443), difference scores are “perhaps the most widely used congruence index . . .” (Edwards, 1994, p. 52).

However, recommendations against the use of gap scores have been largely documented. Cronbach and Furby (1970) asserted that “although the unsuitability of such [difference] scores has been long discussed, they are still employed, even by some otherwise sophisticated investigators” (p. 68).

At first glance, gap scores are not parsimonious, because an additional, third index is computed on top of the two already measured variables (Johns, 1981). In addition, gap scores present several methodological and substantive difficulties that hamper their usefulness. These methodological problems are (a) ambiguity, (b) variance restriction, (c) decreased reliability, (d) spurious correlation, and (e) multicollinearity. The substantive flaws are (f) lack of construct validity, (g) imprecise psychometric properties, (h) psychological constraints, and (i) lack of monotonicity. These problems and alternative approaches are examined below.

Ambiguity

The first methodological problem is conceptual ambiguity, resulting from the apparent “equal but opposite contributions from each component” (Edwards, 1994, p. 53). The symmetry assumption is verified only when the variances of the component scores are homoscedastic (Edwards, 1994). Otherwise, the equal magnitude and opposite signs assumption fails because “the resulting difference will primarily represent the component with the larger variance” (Edwards, 1994, p. 53). Thus, heteroscedasticity clearly violates

the assumption underlying the use of this equal contribution index. An extreme example would be the case where one of the components is a constant. In that case, the “composite score simply represents a rescaled version of the other component” (Edwards, 1995, p. 308). Perhaps Babakus and Boller (1992) encountered a similar phenomenon when they found that perception scores dominated the difference scores, with little contribution from expectations scores. In addition, the observed variability may be subject to a sample effect, which could explain why Babakus and Boller (1992) found instability to prevail in the factor patterns of instruments composed with differences scores. In the same vein, Edwards (1994) mentioned that indices could cause confusion about the effects of their components and conceal the relative contribution of each component towards the outcome.

Variance Restriction

The idea of computing an index is founded on the implicit assumption that it will improve the measurement process, such that the variance explained by the composite would exceed that of the commensurate measures (Edwards, 1994). While the index may be superior to the individual components taken separately, at best, it can only explain as much variance as “both components considered *jointly* . . . and, in most cases, will decrease [the explained variance]” (Edwards, 1994, p. 56). This problem is referred to as variance restriction, which have been characterized by the “ceiling and floor effects” metaphor (Tabachnick & Fidell, 1989, p. 347).

Variance restriction is typical of “motherhood” variables, that is, variables for which more is consistently better than less (Brown et al., 1993, p. 131). This phenomenon is attributed to the fact that expectations (E) normally exceed perceptions (P) (Wall & Payne, 1973), with predominantly negative scores for differences (D). Because E scores are generally larger in magnitude than P scores, the range of D scores for higher P scores will be smaller than that for lower P scores. For example, with seven-point Likert scales, D ranges from 0 to -2 if $E \geq P$ and for $P = 5$. However, for $P = 3$, D ranges from 0 to -4. When this systematic variance restriction situation occurs, the variance of the difference

score decreases with the absolute magnitude of the difference scores and it invalidates the homoscedasticity assumption required by many statistical procedures (Peter, Churchill, & Brown, 1993).

PZB acknowledged that the variance of the difference scores may be restricted in the upper range of quality scores (Parasuraman et al., 1993). However, they countered that it is irrelevant when the scores are used for diagnostic purpose, one of its principal features. They admitted that it might hamper multivariate analyses of the data. They added that the problem will be “serious only if the difference-score measure is the dependent variable” and that some statistical techniques, such as generalized least-square regression as well as transformation of variables, are available to counter this problem (Parasuraman et al., 1993, p. 143). Furthermore, PZB insisted that the quality gap scores were generally used as the predictors, not outcomes.

Nunnally (1967) stated that “the larger the variance of the linear combination [of variables], the more the reliability” (p. 231). Hence, reducing the variance of the composite, as discussed *supra*, will *de facto* result in reducing its reliability, an issue that is discussed separately next.

Decreased Reliability

According to Johns (1981), gap scores generally reduce the reliability of the difference score as compared to average reliability of its components. The reliability of an algebraic difference can be computed using the equation suggested by Lord (1963, p. 32) and substituting sample variance (s^2) to the population variance (σ^2) when necessary

$$r_{\text{diff}} = \frac{\sigma_1^2 r_{11} + \sigma_2^2 r_{22} - 2r_{12}\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2r_{12}\sigma_1\sigma_2}$$

where r_{diff} is the reliability of the difference score, r_{11} is the reliability of the first component, r_{22} is the reliability of the second component, r_{12} is the correlation between the

components, σ^2_1 is the variance of the first component, and σ^2_2 is the variance of the second component.

When the correlation between the components (r_{12}) is equal to zero—when there is no correlation—the reliability of the difference score is a weighted average of its components' reliabilities. Not only there is no gain, but also the absence of correlation between the components is a bad symptom of the pertinence of the index. Johns (1981) stated that it “would call into question the [convergent] validity of the components themselves, raising doubts as to whether they refer to the same stimulus object” (p. 448).

When the correlation is positive, a very likely event, attenuation occurs and the reliability of the difference score will be less than the weighted average of its components' reliabilities (Johns, 1981). In fact, positive correlation between the components is actually expected from measures originating from the same source (Lord, 1963), such as SERVQUAL. Consequently, the contribution of error of measurement to the index is overestimated (Edwards, 1995). When negative—inverse—correlation is encountered, fortunately an unlikely prospect, the reliability “will be magnified” (Johns, 1981, p. 448). As summarized by Nunnally (1967), “the pattern . . . of signs in the linear combination [of variables] has a direct effect on the reliability of the combination (p. 231).

The sign of the correlation score can be manipulated by reversing the coding of any one variable, such that the researcher can set the sign of the correlation coefficient, but not its magnitude. But the researcher can influence the magnitude of a reliability coefficient through the sign of the correlation between components. A negative correlation entails adding the same constant to the numerator and the denominator of a fraction and always results in a new fraction larger than the original, while a negatively signed constant will always result in a smaller new fraction. It thus follows that underestimating or overestimating the reliability with the use of gap scores becomes a method effect. In the previously mentioned studies correlation between SERVQUAL factors is always reported in a positive fashion.

Peter et al. (1993) noted that the scale reliabilities reported by PZB (1988) were computed using the Cronbach's α formula instead of the formula indicated by Lord (1963). Also, in studies of SERVQUAL where this issue was not addressed, there is a possibility that the reported reliabilities underscored the average reliability of the components.

In the case of a causal research, the attenuation present weakens the relationship to the dependent variable, defeating the intended purpose of the study (Peter et al., 1993). In addition, the same authors cautioned that it may not be proper to use lower reliability as evidence of discriminant validity when, in fact, the measures correlate at a higher level and possibly belong to the same construct (Peter et al., 1993).

Spurious Correlation

Difference scores are "systematically correlated with their component parts . . . even when the components scores were generated randomly" (Johns, 1981, p. 451). The same author warned that there might be spurious correlation between the gap score and other variables that were legitimately related to the components. This spuriousness is often "masked by the face validity of the difference scores" (Johns, 1981, p. 451). This may have been the case when Babakus and Boller (1992) examined the relationship between the index and some criterion validating variables.

In any case, the use of difference scores always creates "an artificially negative correlation between expectations and [differences]" (Olivier, 1981), when controlling for perceptions. This has been designated as the "*logical constraint*" (Wall & Payne, 1973, p. 322).

This phenomenon can best be illustrated by an example. Using the familiar $D = P - E$ model, and assuming two events receive the same P score of 5, an E score of 7 will make $D = -2$, while an E score of 1 will result in $D = 4$. Thus, lower expectations will entail higher differences, with negative correlation between these variables. Consequently, any other variable highly correlated with E will have a negative correlation with D, for no other reason. *Contra*, when D is computed as $E - P$, the inverse relationship is found

between perceptions and differences (Wall & Payne, 1973). It must also be noted that in both cases the variable that is not negatively correlated will be positively correlated, such that using the $D = P - E$ formula, D and P would be positively correlated, when controlling for E. Consequently, any independent variable that has a positive relationship with P will also demonstrate a positive relationship with D. In a reciprocal fashion, D and E will be negatively correlated, holding P constant (Teas, 1993).

Multicollinearity

The levels of the difference scores are not independent from their components (Johns, 1981) because their sign is directly related to the level of one component, when controlling for the other. So, quality can be expressed as a linear combination of either expectations or perceptions, holding the other component constant (Teas, 1993). This creates excessively difficult conditions to argue that the “difference score is measuring something unique from the component variable” (Peter et al., 1993, p. 660). Not surprisingly, multicollinearity problems have been reported with the use of difference scores (Olivier, 1981). In addition, the dependency between the components can be attributed to a method effect because the respondent uses the same scale, merely varying his perspective towards the issue at hand between expectations and perceptions (Peter et al., 1993). The subsequent sub-sections will address the substantive problems encountered when using gap scores.

Lack of Construct Validity

The difference score ought to demonstrate through rigorous convergent and discriminant validity tests that it “*is discriminable from the variables measured by its components*” (Wylie (1974), as cited in Johns, 1981, p. 453). Indeed, in addition to being deeply rooted in theory, the construct validity of the difference score must be proven distinctly from that of the components. This test is failed every time one of the components is a better predictor of a criterion. Although PZB presented compelling evidence that the scale fulfilled the previously stated requirements (Parasuraman et al., 1988), the criterion

test was failed in the study by Babakus and Boller (1992). In addition, the distinctness of the gap score was seriously challenged by the option to express the difference score as a linear combination of at least one of its components (Brown et al., 1993).

Discriminant validity is verified empirically by low observed correlation between the factors (Anderson & Gerbing, 1988). As previously demonstrated, the above mentioned depressed reliabilities will attenuate the correlation between the measures of the different constructs, creating “the illusion of meeting discriminant validity standards simply because of low reliability” (Peter et al., 1993, p. 659). In other words, discriminant validity is achieved only because the scale is unreliable (Brown et al., 1993). On this issue, PZB’s position was that a strong correlation between the scores and their components is not an obstacle to discriminant validity (Parasuraman et al., 1993). They asserted that because quality is a function of the discrepancy between the components (Parasuraman et al., 1985), there is an explicit relationship between them, which will take the form of a high correlation.

Because discriminant validity implies a low correlation between supposedly different constructs (Campbell, 1960), and because quality and its components are not unrelated, PZB claimed that “inferring poor discriminant validity . . . on the basis of . . . correlation [of the gap score] with its components is inconsistent with the definition of discriminant validity and hence inappropriate” (Parasuraman et al., 1993, p. 143).

Imprecise Psychometric Properties

According to Cronbach and Fury (1970), computing a difference score arbitrarily attributes a weight of one to the relationship between two variables, forbidding the data to reveal itself through an endogenous parameter. Furthermore, they added that “there is little reason to believe and much empirical reason to disbelieve the contention that some arbitrarily weighted function of two variables will properly define a construct” (Cronbach & Furby, 1970, p. 79). Carman (1990) used this lack of scholarly knowledge about the psychometric properties of gap scores to justify his position of not using them in his study.

Psychological Constraints

Psychological constraints are related to the fact that the expectations generally exceed the perceptions. This topic has been considered indirectly *supra* in the context of variance restriction. This particular problem is connected with the economic concept of non-satiation in consumer preference theory, positing that *more is better than less* (Eaton & Eaton, 1995). In surveys, most respondents will “rarely state that there should be (E) less than there is (P)” (Wall & Payne, 1973, p. 323). Following this line of thought, it may be asserted that the SERVQUAL items are “loaded”, inconspicuously influencing the answers, and introducing a “systematic bias” in the results (Alreck & Settle, 1995, p. 98-99).

However, the evidence revealed that there is variability in the level of the scores (Parasuraman et al., 1991a), reducing the strength of the assertion that the scores are skewed towards a single direction. Furthermore, Tan and Barnes (1996) encountered perceptions exceeding expectations on some dimensions. Using a modified SERVQUAL instrument, these authors found positive gaps for 13 pairs of items, no difference for 2 and negative gaps for 8. Thus, perceptions exceeding expectations were not rare occurrences, but were observed for the majority of computed gaps. Hence, the evidence does not always support the assertion that the psychological constraint is an impediment to using gap scores.

Lack of Monotonicity

Stipak (1979a) saw a lack of monotonicity problem with the computation of gaps between perceived delivery and expectations. He posited that the use of subjective measures might be appropriate if “an increasing monotonic relationship exists between the measure and the underlying quality dimension . . . [However,] at the individual-level, monotonicity will not generally occur for respondents for whom a discrepancy process of evaluation applies . . . [, unless] the ideal, best performance” (p. 424) standard is used. The underlying rationale is that the gap varies inversely to the performance in a consistent fashion against the ideal standard, assuring monotonicity. In his opinion however, the

rational citizen weights marginal costs and benefits when setting individual standards, and “evaluations will begin to decrease at some point with further increases in performance” (p. 424), which violates the monotonicity assumption.

It should be noted that the SERVQUAL scale uses the paradigm of ideal expectations, where “the service feature being assessed is a *vector attribute*—that is, one on which a customers’ ideal point is at an infinite level” (Parasuraman et al., 1994b, p. 116). When examining this issue, Teas (1993) concurred that ideal expectations “reflect constantly increasing levels of quality” (p. 19).

Teas (1993) proposed the use of the absolute difference scores approach, instead of the algebraic difference scores. However, because absolute difference scores are only a variation of the basic gap scores indices, they suffer the same problems as standard gap scores (Edwards, 1994). This approach presents no significant methodological improvement, instead reducing the diagnostic value of the scores because these coefficients are unidirectional. Furthermore, the absolute difference scores are not monotonic.

Alternative Approaches

Although the unanimous verdict on gap scores is not favourable, there are a few options open to the researcher. The first alternative would be to assess empirically the appropriateness of using the difference scores. Not taking error of measurement into account, Edwards (1994) suggested that this can be done by testing both components as separate predictors of some outcome variable and verifying that they show effects that are independent, significant, and “equal in magnitude but opposite in sign” (Edwards, 1994, p. 57).

However, several empirical investigations of the SERVQUAL scale evidenced that the perceptions battery alone was as good or better than gap scores as a predictor of some outcome variable (Babakus & Boller, 1992; Carman, 1990; Cronin & Taylor, 1992), but not in all cases (Brown et al., 1993). In fact, PZB conceded that even in their own studies, perceptions alone were a better predictor of some criterion variable than SERVQUAL

(Parasuraman et al., 1993, p. 145). They countered, though, that the gap scores provided *richer* information than the perceptions battery alone, because they indicated the problem areas more accurately. Indeed, two items may obtain the same score on the perceptions scale, implying identical results, and only interaction with the expectations scale will reveal the direction and the importance of a quality gap. They added that only separate measures allow for a longitudinal tracking of the dynamic evolution of customers' attitudes. They concluded that collecting the perceptions and expectations measures distinctly allows the versatility to use the measures separately or jointly, depending on the specific needs of the analysis.

A second alternative would be to use direct measures, that is, variables that "elicit a relative comparison between two components" (Edwards, 1994, p. 89), through a single value. Such measures are certainly more parsimonious (Johns, 1981) and less burdensome on respondents who are only confronted with half of the number of items (Peter et al., 1993). This approach to modifying SERVQUAL has been proposed (Carman, 1990) and experimented (Brown et al., 1993; Parasuraman et al., 1994a).

On the issue of direct measures being free of the problems that are typical to difference scores, the evidence is mixed. On one side, direct measures are seen as having better psychometric properties and yielding better quality empirical results (Peter et al., 1993). However, Edwards (1994) asserted that the task to "implicitly or explicitly calculate a difference in the process of generating a response" has merely been passed over to the respondent, making the measures "susceptible to the same problems associated with difference scores" (p. 90).

Furthermore, a direct measure compounds the problem of the relative contribution of each component, which cannot be singled out any more, nor can the symmetry assumption be verified (Edwards, 1994). PZB cautioned that using direct measures also implies the loss of the potential diagnostic value of measuring components separately (Parasuraman et al., 1994b). In addition, when direct measures are used with SERVQUAL

items, they “persistently tend to overstate customers’ assessments” (Parasuraman et al., 1994a, p. 35).

A third possible solution would be to use both batteries with compatible multivariate analysis methods. One of the techniques suggested by Peter et al. (1993) is the hierarchical regression statistical procedure, focussing on the incremental variance accounted for by the sequential addition of one component after the other. Another approach they mentioned is simultaneous estimation through Structural Equation Modelling (*SEM*) (Peter et al., 1993).

Notwithstanding the elements covered in the above discussion, Boulding et al. (1993) concluded that the gaps approach underlying the SERVQUAL scale remains the usual manner in which the scale is operated.

Summary of the Issues with the SERVQUAL Scale

The various issues that were raised about either version of the SERVQUAL scale were examined. The issue of negatively worded items had been resolved in the 1991 version of the instrument, but the other five issues are still under scrutiny.

Although conceptually sound, the multidimensionality of the scale remains a contentious issue, perhaps in view of methodological problems. The universality of the instrument received mixed evidence and is still questioned. The distinct measure of the relative importance of the dimensions appears needed and may contribute to enlighten the universality issue. Also, the evidence demonstrates that a distinct measure of expectations is warranted, in view of the unique information it carries. Finally, evidence against the use of difference scores was reviewed and alternative methodologies have been proposed.

Applications of the SERVQUAL Scale to Municipal Services

Crompton and Mackay (1989)

Crompton and Mackay (1989) made a limited adaptation of the SERVQUAL scale for use with municipal services, as they “sought to identify the relative importance of these

[five] dimensions in their contribution to perceived service quality in selected recreation services” (p. 369) for municipal recreation programs. Data were collected through an in-person, self-administered questionnaire to a stratified random sample of participants to municipal recreation programs in a regional area comprising five towns in Eastern Canada. The response rate exceeded 95%.

Contrary to PZB who had elicited individual importance score for each dimension (Berry et al., 1988), Crompton and Mackay (1989) obtained direct comparative measures. The instrument consisted only of a comparative rating scale to rank order the five dimensions of SERVQUAL by importance. A constant-sum scale was used, eliciting the allocation of 100 points among the five dimensions. This procedure was adopted by PZB for the refined version of SERVQUAL (Zeithaml et al., 1990). Crompton and Mackay (1989) mentioned that the use of the constant-sum scale aimed at reducing the halo effect that may prevail with Likert-type or semantic differential scales, for which respondents have a propensity to rate most or all items equally. The obtained scores for each dimension were scaled in a manner such that the highest score in magnitude became a rating of 100 and the other scores “were adjusted proportionately. This procedure gave the data the basic requirement of equivalency across subjects” (Crompton & Mackay, 1989, p. 371). In addition, Crompton and Mackay (1989) standardized the scale. However, this procedure has not been reproduced in subsequent studies by other researchers.

Crompton and Mackay (1989) found that age, gender, or income of respondents did not affect the importance ratings. They used analysis of variance and Duncan’s multiple range test to assess the hypotheses that the relative importance of dimensions would vary between participants in a program, and between programs. Crompton and Mackay (1989) reported significant differences in the relative importance of the dimensions within programs, with the Reliability dimension consistently the highest-ranking (Crompton & Mackay, 1989). Their findings regarding between program differences were mixed although some differences appeared to prevail. They hypothesized that the source of the

variability was some attribute of the service. They classified services along two axis, staff involvement and dependence on facilities. However, the hypothesis failed to provide a coherent explanation of the variations. Again, the Reliability dimension was the highest-ranking in all quadrants (Crompton & Mackay, 1989).

Carlson and Schwarz (1995)

Carlson and Schwarz (1995) adapted the SERVQUAL scale to assess the quality of service for a county (henceforth referred to as the County scale). Based on information gathered in focus groups of employees and citizens, the authors conceptualized eight dimensions to quality for public services, which comprised 35 paired-items, including 16 from SERVQUAL and 19 new items. The only SERVQUAL construct maintained was Reliability, along with three of its items. The other items were classified under seven new dimensions: (a) *convenience*, (b) *security*, (c) *personal attention*, (d) *problem-solving approach*, (e) *fairness*, (f) *fiscal responsibility*, and (g) *citizen's influence*.

The average score of the items comprised in each dimension were used as index scores. The quality gap was measured by comparing expected to perceived levels of service, where the expected level of service was set as the individual's satisfaction threshold. Carlson and Schwarz's (1995) major finding was that the expectations were not anchored at the perfection level—the top of the scale—but were closer to 4 on a 1-to-5 scale, with one dimension at 3.7, four at 4.0 and three at 4.1. The largest gaps were observed with the (a) fiscal responsibility, (b) fairness, and (c) citizen influence dimensions, in decreasing order of magnitude. The dimensions that were rated as the most important were, also in descending order, (a) fairness, (b) fiscal responsibility, and (c) convenience. This pointed conspicuously to fairness and fiscal responsibility, which were both very important to the residents and were very deficient in their assessment. None of the customary scale indices were reported.

Carlson and Schwarz's (1995) study again raises two previously considered methodological issues, the dimensionality of the scale and the use of gap scores.

Concerning the dimensionality, no evidence was given that the hypothesized relationships between the indicators and the factors in the scale were assessed through a technique amiable to latent variables, such as factor analysis. In addition, Carlson and Schwarz (1995) mentioned that they used the gap scores for their diagnostic power.

One last finding of this study is that “employees consistently underestimated citizen’s perceptions of current service quality” (Carlson & Schwarz, 1995, p. 32). This belief was attributed to the general predominance of complaints over positive events in feedback given to employees.

Donnelly et al. (1995)

The last preceding result was corroborated in another study using SERVQUAL. Donnelly et al. (1995) found that local governments’ managers frequently over-estimate customer expectations. The authors of the study also argued in favour of the ability of the scale to serve as a diagnostic tool. However, no results were made available for scrutiny.

Summary

The above studies are the only ones that have been traced utilizing the SERVQUAL scale to assess municipal services. The paucity of studies in this area is an indication of the importance of the present study. However, the scale has found use in the assessment of many other public services, such as higher education (Ford et al., 1993; Kettinger & Lee, 1994; Moore, 1994; Zeilanga, 1994), health care (Quiram, 1995; Whitman-Smithe, 1995), military (Campbell, 1994; Orwig, 1994), and museums (Tan & Barnes, 1996).

These studies also raise several previously mentioned difficulties concerning SERVQUAL, but its conceptual underpinnings appear undisputed. Thus, SERVQUAL remains a powerful and popular instrument to assess perceived quality for public, if not municipal, services, as more and more recent studies would indicate.

Measures of Satisfaction and Relationship to Quality

This last sub-section is devoted to the measurement and modelling of satisfaction. This will entail an examination of the practices in measuring satisfaction for local services. Following is an examination of how satisfaction and quality differ and how they can be linked in a causal relationship.

Measuring Satisfaction with Municipal Services

Satisfaction with municipal services has been measured in many different ways. In studies using archive data, satisfaction was measured through surrogate measures, such as complaints and incidence of specific events. These indicators did not appear to be effective indicators of the citizens' opinions, because "reported crime rates and citizen complaints tend to underestimate dissatisfaction" (Hatry & Blair, 1976, p. 131).

In studies targeted for residents' reply, more conspicuous items were used in the questionnaires. In some of the earlier studies, quality and satisfaction were not always measured as distinct constructs. Pelissero (1978) used a five-point scale to rate seven services on their quality, ranging from *very good* to *poor*, and labelled this measure as the "General Service Satisfaction". This approach was also used in the seminal work by Stipak (1974), in which the difference between the two concepts was also minimized.

For commercial services, Spreng and Singh (1993) suggested that the satisfaction items be directed to the firms, not the service. However, other studies aimed satisfaction and other outcome items at the service, not the firm (e.g., Teas, 1993; Waldman & Gopalakrishnan, 1996). For towns, however, many sources of satisfaction or dissatisfaction, such as employment opportunity and geographic location, are unrelated to services. Consequently, satisfaction items pointing to the service may avoid confusion.

Studies used one or more items to measure satisfaction. Depending on the study, the satisfaction construct comprised (a) single measures, only one item; (b) multiple

measures, a collection of one-item scales addressing multiple services, and (c) composite measures, a cluster of assorted items meant to represent the concept of satisfaction. Studies falling in these three categories are reviewed below.

Single Measures

Satisfaction has been measured directly through a single, overall item eliciting the satisfaction with a particular service (Goodsell, 1981), or services in general (Marans & Rodgers, 1975). The measure of satisfaction has also consisted of an indirect measure items such as *getting your money's worth* (Lovrich & Taylor, 1976; Roth et al., 1990).

According to Hayes (1991) "the observed score on their one-item measure is not highly related to the actual levels of satisfaction" (p. 44). Because the internal consistency of a scale increases through the use of multi-item scales (Hayes, 1991), multi-item scales are preferred to single-item scales.

Another weakness with such one-item measures is the confusion that arises from the potentially divergent interpretations of a vague item. On this topic, Stipak (1979b) posited that "a vague survey item asking citizens how satisfied they are with police services may reflect a wide variety of factors, such as whether the respondent was (a) a victim of a crime recently; (b) stopped by local police recently, and if he was stopped, (c) how the police treated him" (p. 51).

Multiple Measures

Satisfaction has also been elicited through multi-item scales composed of several items, each targeting an individual service, such as police, fire, garbage collection, etc. Murdoch and Schriener (1979) had items surveying satisfaction on nine municipal services, as well as other non-local services. Taylor (1982) used a seven-point scale assessing satisfaction with police, fire, library and public park services. Das et al.'s (1995) study comprised items eliciting responses on satisfaction with 18 typical municipal services. In those instances, the individual services were analyzed individually.

Schuman and Gruenberg (1972) used a four-item scale to measure satisfaction with a corresponding number of municipal services. They found large discrepancies between the individual service satisfaction levels. This *differentiation* lead them to rule out the hypothesis that this “set” of items were indicators of an unidimensional “underlying construct” (Schuman & Gruenberg, 1972, p. 373) of a generalized satisfaction with municipal services.

Using a slightly different approach, Brown and Coulter (1983) used a multiple-item convergent measure. Several items, mapping their view of the content of the satisfaction construct, were used concerning a single service. In the Brown and Coulter (1983) study, answers on a four-point ordinal scale were elicited on three items: (a) the general satisfaction with police protection, (b) the response time, and (c) the treatment by police. Nonetheless, the dependent variables were analyzed independently. Although the scale covered several aspects of the satisfaction concept, it did not benefit fully from its structure, as composite measures would.

In conclusion, multiple measures do not represent a significant improvement over single measures. Hence, the caveat of single-item scales also applies to multiple measures.

Composite Measures

Indices formed with multiple items representing complementary aspects of the satisfaction concept domain were also used. Measuring satisfaction with police services, Brede (1985) used an additive index comprising four items, namely (a) overall satisfaction, (b) willingness to report a crime, (c) perceived reliability, which can be construed as a quality indicator, and (d) fairness of treatment. Lyons et al. (1992) used a two-item index, overall satisfaction and overall doing a good job, which can also be viewed as a quality indicator.

Additive composite measures comprising semantic differential scales to measure satisfaction have been used. McClendon and O'Brien's (1988) study used four items measured on a seven-point scale, for which they reported a reliability score of .90. Also

forming a composite measure, but in a non-additive manner, Percy (1986) used a four-item satisfaction scale. In his scale, three specific assessment items were tested as determinants to the fourth item, which covered the general assessment. He found that two out of the three specific items were significant predictors of the overall measure.

Andreassen (1994) examined satisfaction as a latent variable, with four items as indicators, measured on a six-point linear, numeric scale. One item was an overall measure, another one was service specific, and the other two were disconfirmation assessments. The unidimensionality of the factor was ascertained.

Distinguishing Satisfaction from Quality

The concepts of quality and satisfaction are very close and, according to Iacobucci, Ostrom, and Grayson (1995), they are sometimes used as synonyms by researchers and practitioners. This is not surprising because both concepts have been modelled as discrepancies between expectations and perceptions towards some service or product.

According to Dabholkar (1993), conceptualizing and operationally setting satisfaction and quality distinctly appears more important to academics than practitioners who have an interest in the concepts only for their predictive power of the customer's behavioural intent and who often use the same items when measuring satisfaction and quality. However, "managers should recognize the importance of the relationship between perceived quality and satisfaction" (Gotlieb, Grewal, & Brown, 1994, p. 884) and it may be useful for them to distinguish between these two concepts.

According to Olivier (1993), there are four levels on which quality and satisfaction differ conceptually. These levels are (a) underlying dimensions, (b) focus of expectations, (c) requirement of experience, and (d) underlying antecedents. These conceptual issues are examined subsequently, along with (e) the empirical perspective.

Underlying Dimensions

First, according to Olivier (1993), the quality judgement is contingent upon a limited set of dimensions, which are all related to the service act, whereas the satisfaction

judgement is based on a larger set of dimensions, some of them not related to the service experience itself. For instance, although a high quality judgement can be formed about a dining experience, a problem with a credit-card authorization can induce dissatisfaction. This would concur with Crompton and Mackay (1989) who posited that “satisfaction is a psychological outcome emerging from an experience, whereas service quality is concerned with the attributes of the service itself” (p. 368).

Focus of Expectations

Second, Olivier (1993) also posited that while quality sets expectations at the normative level, satisfaction judgements may incorporate other expectations determinants, such as desires, needs, and fairness. Thus, the satisfaction judgement is contingent upon the situation. Based on their research, PZB mentioned that “respondents gave several illustrations of instances when they were satisfied with a specific service but did not feel the service was of high quality” (Parasuraman, Zeithaml, & Berry, 1988, p. 16).

Requirement of Experience

Third, Olivier (1993) mentioned that satisfaction judgements can only be formed as a result of experience, while consumers can have a perception of quality towards a service without experiencing it, such as high quality for a renowned restaurant. This difference, however, appears to blur the conceptual distinction that was previously established between quality, which it aims to qualify, and its expectations component, which it actually qualifies. This difference does not apply to SERVQUAL, which is experience-driven.

Another similar distinction between quality and satisfaction has been made on the basis of their relationship to experience. According to Boulding et al. (1993), in the satisfaction literature the emphasis is on perceptions related to the most recent, specific transaction, which are specific experience-dependent, whereas the quality literature is more concerned with cumulative perceptions, which are also experience-dependent although not specific. However, according to Dabholkar (1993) this is not a valid threshold upon which to segregate the two concepts, because an available body of dissenting literature indicates

“that . . . customer satisfaction and service quality can both be conceptualized in terms of a given experience or in global terms” (p. 12). Thus, the transactional versus global argument may then be more symptomatic of the quality and satisfaction literatures evolving independently than from substantive conceptual or empirical distinctions.

Underlying Antecedents

Finally, Olivier (1993) contended that a fourth distinction is that there is only a limited set of antecedents to quality, most of them communication dependent, whereas satisfaction is influenced by a wider set of “cognitive and affective processes including equity, attribution and emotion” (p. 77). This argument is close to the one on underlying dimensions, which was examined *supra*.

Empirical Perspective

Although most of the distinction efforts have examined the issue from a conceptual perspective (Iacobucci et al., 1995), some empirically focussed studies have also been conducted. When differences were empirically observed, the emergence of distinct factors for the concepts stemmed mainly from the measures employed (Dabholkar, 1993).

Dabholkar (1993) added that if both constructs are modelled in a similar manner—both as gaps between expectations and perceptions, both at the transactional level, or both at the global level—a single factor is likely to emerge (Dabholkar, 1993). In addition, Dabholkar (1993) posited that if the constructs were wilfully modelled differently “an artificial separation of the factors” occurs, although they “are likely to be highly correlated” (p. 14).

Olivier (1993) took exception to the proposition that the concepts vary in unison and posited that measurements of the factors “at variance”—dissatisfaction with high quality or satisfaction with low quality—are conceivable (p. 78). He stated that the concepts are neither independent nor enjoy a one-to-one positive relationship, but rather “that they are only modestly related” (Olivier, 1993, p. 78). Taking a similar position, Mason (1996, p. 76) added that “ratings of customer service are notorious in their ‘inter-rater variability’

with some respondents providing a high score and others a low score to what is objectively the same level of service”.

The distinction between quality and satisfaction has been investigated empirically in the municipal service context. Lyons et al. (1992) compared services between jurisdictions using a perceptual measure of service quality. Their assessment revealed that quality and satisfaction were significantly distinct concepts in spite of the significant correlation between them.

Linking Quality and Satisfaction

While the above academic debate is still on going, there is a very strong intuitive crunch that quality and satisfaction coalesce. Considering both concepts in the same context, some researchers have hypothesized that satisfaction was an antecedent of quality, while others saw the influence exercised in the reverse order.

Fuelling the satisfaction-causes-quality approach, Olivier (1981) posited that satisfaction decays into an attitude towards the product or service, and therefore precedes it. This resulting attitude is perceived quality (Gotlieb et al., 1994). This position has received some support from the conceptual-only perspective (Swartz & Brown, 1989). Taking the opposite quality-causes-satisfaction position, Iacobucci et al. (1995) posited to the contrary that properly customer rooted “quality improvements should lead to customer satisfaction” (p. 296).

The recent emergence of SERVQUAL as an effective tool to measure quality has instrumented researchers to pursue the empirical examination of the question. Lately, several studies have attempted to establish a link between service quality, as measured by SERVQUAL or other means, and satisfaction. In these empirical studies, both concepts have been examined alternately as the possible antecedent of the other. The bulk of the evidence is “that service quality is an antecedent of consumer satisfaction” (Kettinger & Lee, 1994, p. 743).

Building on Yokoyama (1992), quality can be modelled as an antecedent to satisfaction in three alternate ways: (a) the *perception-only* model, (b) the *discrepancy* model, and (c) the *simultaneous* model. According to Yokoyama (1992), the discrepancy model contends that *gap-measured* quality influences satisfaction while the perception-only model, which he labelled *performance-based change model*, states that perceptions quality alone influences satisfaction. Some studies verifying these models are reviewed below.

Perception-Only Model

Along the perception-only influence satisfaction canvas, the previously reported Cronin and Taylor (1992) study examined an array of models, hypothesizing alternate causal structures, including a reciprocal one, in an attempt to causally link quality to satisfaction. Their SERVPERF instrument outperformed SERVQUAL because it attained an acceptable fit ($AGFI \geq .90$) in all four samples using the *SEM* technique. Satisfaction was measured by a single item that elicited the respondents' attitude towards the service, not the firm itself. They found that the quality as an antecedent to satisfaction link was significant, while the reciprocal was not.

Gotlieb et al. (1994) also found that only the quality-causes-satisfaction link was significant. They used a ten items reduced SERVQUAL perception-only scale, which they found to be unidimensional. The satisfaction scale comprised three items: (a) happiness about the decision, (b) did the right thing, and (c) overall satisfaction. Gotlieb et al. (1994) explicitly rejected the inverse, as well as the reciprocal, possible relationships between quality and satisfaction. Expectations were measured indirectly through disconfirmations items, with the scale ranging from *better than expected* to *worse than expected* on focal dimensions, such as courtesy, efficiency, and technical care, and contextual dimensions, like room and meals. This study was not classified under the quality as gap category, because expectations were not considered from the ideal standpoint and, more importantly, because disconfirmations were seen as an antecedent to, and not part of, the quality measures.

Discrepancy Model

Parasuraman et al. (1991a) examined the relationship between SERVQUAL gap scores and an item eliciting the respondents' propensity to recommend the firm. Through a test for difference of group means, they found that higher quality scores were related to a higher level of satisfaction. Although *propensity to recommend* may seem to be only an indirect measure of satisfaction, it is used for the formation of composite measures of satisfaction in other studies (Larsen, Attkisson, Hargreaves, & Nguyen, 1979; Rosen & Karwan, 1994).

Another study by Kettinger and Lee (1994) examined quality and satisfaction with the Information Services function in a college. In this study, each individual SERVQUAL dimensions, as well as other independent variables, were assessed as a predictor of satisfaction. Satisfaction was measured through a four-item scale. Using a regression, Kettinger and Lee (1994) found that the Reliability and Empathy dimensions both stood out significantly as determinants of satisfaction.

In summary, both the perception-only quality and the gap-measured quality have been used with some success as determinants of satisfaction, but the bulk of the evidence is in favour of the perception-only approach.

Simultaneous Model

Peter (1993) raised the possibility of a third, simultaneous, modelling approach. He proposed that expectations and perceptions be measured separately and modelled distinctly, as they would both influence satisfaction. It is possible that this approach could be tested using *SEM*, a technique that has the capacity to represent complex relationships between variables. However, no studies empirically exploring this approach have been found.

Summary of the Relevant Research

Different studies that examined several potential determinants to satisfaction with local services have been reviewed. None was found to produce compelling evidence on

what really causes this satisfaction, although perceptual measures appeared to be in the right direction.

The issues concerning quality of service were reviewed next. For measurement purpose, local services were found to be similar to commercial services. Consequently, valid instruments from Marketing can be safely used in the municipal context for assessing the quality of service.

The SERVQUAL (Parasuraman et al., 1991a) instrument has been examined. Its development was traced, some issues were addressed, and applications to the municipal service context were reviewed. The instrument was found to be suitable for use with local services.

Finally, the measurement and modelling of satisfaction were examined. Alternative manners of measuring satisfaction with municipal services were assessed. The distinction between satisfaction and quality was considered. Also examined was the manner to link them causally.

Conceptual Framework and Research Questions

In this sub-section, we first examine the objective of this research. Then, the specific research questions for this study are developed, based on the theoretical underlaying of quality and satisfaction, as well as the relationship between these concepts in the municipal setting.

Objectives of the Research

This study pursues four aims. The aims are (a) to develop a reliable and valid instrument to determine the quality of municipal services as perceived by residents, (b) to assess the relative importance of the underlying dimensions in the shaping of the quality construct, (c) to develop a reliable and valid instrument to measure residents' satisfaction with the services provided by a municipal government, and (d) to study the relationship

between measures of satisfaction and quality of service offered by a municipality to its residents.

Municipal services are examined through their impact on two market assessment mechanisms, quality and satisfaction. To achieve this, municipal services must be qualified in terms of quality and satisfaction, with a possible relationship between these concepts. This contention is first supported by the ample body of literature examined in the previous section which testified that not only municipal services can be expressed in terms of quality and satisfaction, but that both concepts have been isolated with some success. Secondly, the possibility of a relationship appears likely in view of the fact that a relationship has been determined to exist in other contexts.

The conceptual definitions of the two concepts of interest in the current research—quality and satisfaction—are presented in the following sub-sections and alternative forms of the relationship between them, as suggested by the literature, are considered.

Construct of Quality

Different approaches to defining perceived quality have been examined in the previous section. The model proposed by Parasuraman et al. (1985) has emerged because it presents significant advantages over the other models. The SERVQUAL scale is characterized by a contemporaneous and separate measurements of expectations and perceptions. Carlson and Schwarz (1995) have found that “citizens can distinguish between the quality of service they expect . . . and the quality of service they are actually receiving” (p. 28). This conclusion suggests that the SERVQUAL instrument may be suitable for this study. This avenue is reinforced by the strong existing similitude between municipal and private services and by the successful previous uses of the instrument in studies on municipal services.

In addition, the SERVQUAL expectations items are set at the ideal level. By anchoring one of the batteries, the resulting quality scores display an increasing monotonic relationship between the indicator and the underlying latent variable. Stipak (1979a) has

asserted that when this condition is fulfilled, “the subjective measures can be considered an order-preserving transformation of a true measure of quality, and rankings will be invariant, subject to measurement error in the subjective measures” (p. 422).

SERVQUAL is also characterized by a multidimensional construct of quality. A relationship likely exists between the five dimensions of SERVQUAL in that they each specify a distinct, but contributing, element of the quality concept. Thus, the multidimensionality of the quality construct could possibly be represented through the use of a second-order factor model. This approach not only implies a sustained relationship between the five dimensions, but it specifies them as contributors to an encompassing, underlying second-order factor. Figure 2.4. illustrates the second-order factor model for the quality construct.

The purpose of the first research question is to verify the usefulness of the second-order factor quality model to represent residents’ perceived quality in the context of municipal services. This will address the outstanding issues of the dimensionality of SERVQUAL and its universality.

Research Question 1: Do the data support a second-order factor quality model to represent residents’ perception of quality of municipal services?

Relative Importance of the Dimensions

The review of the relevant literature on SERVQUAL suggests that there is no consensus concerning the equality or inequality of the relative importance of the dimensions. It is therefore warranted to examine this issue using the current data set. The purpose of the next research question is thus to assess the relative importance of the quality dimensions.

Research Question 2: Do the data support equal or different relative importance levels for the quality dimensions?

Construct of Satisfaction

The review of the literature has revealed that the construct of overall satisfaction is often operationally defined by a single item, or at best, by a small cluster of converging items. The satisfaction model links a single satisfaction factor to its indicators. The purpose of the next research question is to verify the usefulness of the satisfaction model to represent residents' satisfaction with municipal services.

Research Question 3: Do the data support the satisfaction model to represent residents' satisfaction with municipal services?

Relationship Between Quality and Satisfaction

The review of the relevant literature highlighted the previous use of two models, (a) the discrepancy model and (b) the perception-only model to causally link quality to satisfaction. An alternative link between quality and satisfaction has also been modelled through (c) the simultaneous model, introduced by Peter et al. (1993) but yet to be empirically explored, that links distinctly expectations and perceptions to satisfaction.

The discrepancy model will not be examined in this study in view of the abundance of evidence against such methodology. The perception-only model is a special case of the simultaneous model, with the causal relationships from expectations to perceptions and satisfaction fixed to nil. It will be examined when testing the simultaneous model. The simultaneous model is illustrated in Figure 2.5. as a path diagram. The simultaneous model assumes that quality is a function of both the expectations prior to delivery and the perceptions towards the service at the time of delivery. Expectations and perceptions are not independent, because expectations influence perceptions, and because they both influence satisfaction. Perceptions have a direct effect on satisfaction. Expectations have a direct effect on perceptions and an indirect effect on satisfaction.

The purpose of the last research question is to verify the usefulness of either the simultaneous model or the perception-only model to explain the relationship between quality and satisfaction. This will address the outstanding SERVQUAL issues of measuring expectations and finding an alternative to gap scores.

Research Question 4: Do the data support the simultaneous model or the perception-only model to represent the relationship between quality and satisfaction with municipal services?

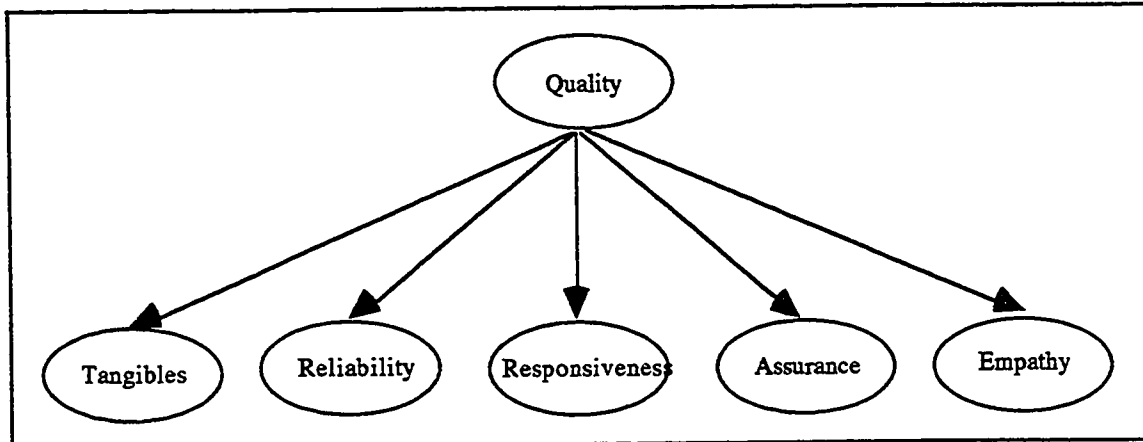


Figure 2.4. A second-order factor model for quality.

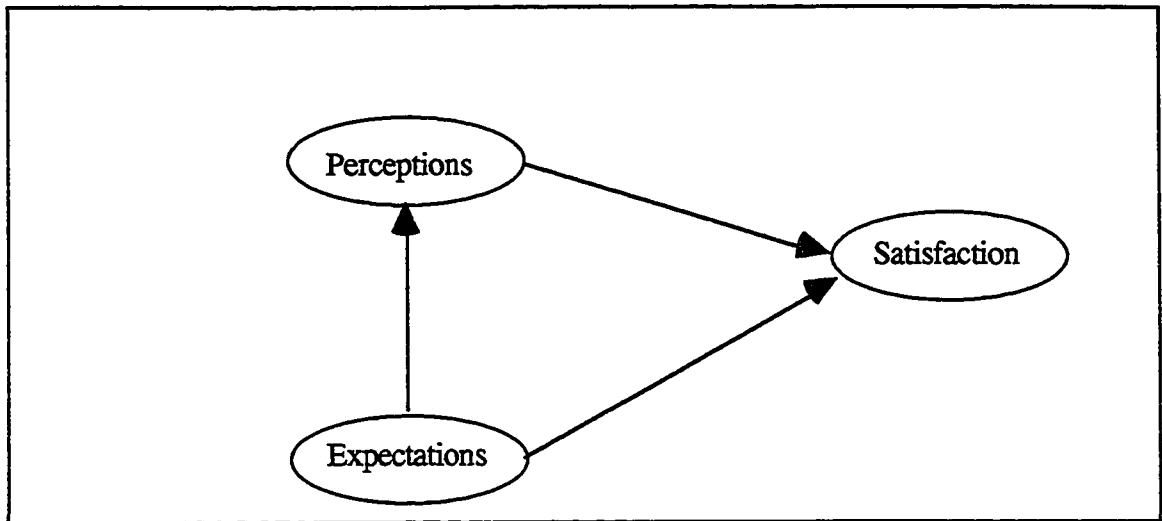


Figure 2.5. Path diagram of the simultaneous model.

METHODOLOGY

In this section, the methodology used for this cross-sectional study will be examined in detail. First, specifications of the instrument design will be presented. Data sources will be examined next. Finally, the statistical tests performed on the data will be discussed.

Instrument Design

The English version of the instrument is included in Appendix B1 and the French version in Appendix B2. The survey was carried out in the Town of Saint-Jean-sur-Richelieu; further details about the study population will be provided subsequently. The survey instrument included a section focussing on a sponsoring organization, items 53 to 57. This section will not be examined because it does not relate to the object of this research.

Independent Measures

The independent variables are items measuring the residents' attitudes towards the quality of the services provided to them by the town. The SERVQUAL instrument was used for this purpose. The SERVQUAL instrument consists of four sub-scales: (a) the expectations battery, (b) the perceptions battery, (c) the relative importance measures, and (d) the overall importance measures. Both importance sections are inserted "between the expectations and perceptions sections" (Zeithaml et al., 1990, p. 175), to prevent confusion when replying to the perceptions battery.

PZB suggested that the instrument "can be adapted or supplemented to fit the characteristics or specific research needs" (Zeithaml et al., 1990, p. 175). Some adaptations of the scale in published studies entail more than minor changes, such as the offspring scale

LODGQUAL (Thompson & Getty, 1994). In this study, the modifications to the scale have been kept to a minimum. This conservative approach when using the instrument in an innovative manner within an unfamiliar setting was suggested by Quiram (1995). Because the use of SERVQUAL in this field of study is still exploratory, the scale was not modified substantially. In addition, this conservative approach was required to enable meaningful comparisons with other replication studies. Nonetheless, to avoid ambiguity, *municipalities* was substituted for *companies* and *residents* for *customers* in the items and instructions. Some items as well as the instructions were abbreviated without altering their meaning.

Dependent Measures

The dependent variables in this study consist of items measuring the satisfaction with municipal services. None of the previously reviewed multi-item measures of satisfaction with municipal services seemed appropriate to be used as-is in this study. Hence, a scale comprising five self-developed or adapted items was elaborated.

The items were designed to assess the municipal services globally; attitudes towards individual services were not solicited for two reasons. First, many municipal services are rarely utilized, such as fire prevention. Eliciting a satisfaction score concerning those services may be futile. It may introduce “acquiescence” (Alreck & Settle, 1995, p. 101) response bias or it may otherwise induce erroneous results. Secondly, because quality had also been examined from a global perspective, approaching satisfaction with a homogeneous framework increased methodological consistency.

The satisfaction items thus tackled attitudes towards municipal services that are assumed to be shared evenly by residents according to their satisfaction level. Each of the five statements aimed at a distinct component of the global satisfaction domain. Because the statements were attitudinal, they were deemed better suited than service specific items to fall within the scope of the “experience and understanding of individual respondents” (Brede, 1985, p. 63). Most of the candidate measures have been used successfully in the past as indicators of satisfaction in other studies. Nonetheless, the ability of the items to

encompass the construct's domain, as well as other validity measures, do require validation and constitute the object of a research question.

The five items consist of statements eliciting responses on seven-level Likert-type scales, anchored only at extremes, with *strongly disagree* matched to value one and *strongly agree* matched to value seven. Seven-point bipolar Likert-type scaling was retained because the number of scaling points is optimal (Alwin & Krosnick, 1991) and because retaining the same scale as the quality batteries raised the level of homogeneity and simplicity of the survey questionnaire.

Minimizing bias appeared an important issue in item construction. Because "posing a satisfaction question in a positive form appears to lead to greater reported satisfaction than posing it in a negative form" (Peterson & Wilson, 1992, p. 65), it was felt preferable not to include only positively worded statements. To offset the yea/nay-saying bias, Alreck and Settle (1995) suggested reflecting "roughly half of the items . . . [such that] half the statements might be negative toward the issue and half positive" (p. 101). In keeping, three positive and two inverted items were included, leading to a balanced approach, free from the assumed bias. When computing the results, the inverted items SAT1 and SAT3 were "reflected such that higher scores indicate higher" (Carmines & Zeller, 1979, p. 68) levels of satisfaction. This procedure aims to achieve an unidirectional scale for more meaningful parameter estimates through factor analysis. These items are reviewed below.

Comparative Rating

The aim of the first item is to elicit the respondents to compare their satisfaction with the services to that of the residents of other towns in the region. Due to the close proximity of the towns and the regional character of print as well as electronic media, it is likely that the respondents possess sufficient knowledge to form an informed judgement on the issue. This item is inverted, with higher scores indicating a lower relative satisfaction with the services.

SAT1: "Persons living in other municipalities of the Haut-Richelieu region are more satisfied with the municipal services they receive than I am."

Source: Self-developed.

Overall Service Satisfaction

The aim of this item is to gather an overall assessment of the residents towards their satisfaction with the municipal services. The statement is worded simply to provide an all-encompassing reply.

SAT2: "I am fully satisfied with the services provided by Saint-Jean."

Source: Based on Item 4.a. "I am satisfied with the services I receive in XYZ" (Teas, 1993, p. 32).

Propensity to Relocate

The purpose of this item is to capture the attitude towards defection in view of poor service. This item is inverted, as higher scores imply a lower level of satisfaction with the services.

SAT3: "Because of the poor quality of municipal services in Saint-Jean, I am thinking of moving to another town."

Source: Based on item "Switch to a competitor if you experience a problem with XYZ's service" of the Behavioural-Intentions Battery (Parasuraman et al., 1994a, p. 45).

Collective Assessment

This measure is included to alleviate the psychological threat that some respondents may face when introspecting their satisfaction, by depersonalizing it. As suggested by Alreck and Settle (1995), the issue is raised to the abstraction level and projected towards others, "referring to the experience or condition of some other person" (p. 102).

SAT4: "Almost all residents are satisfied with the municipal services of Saint-Jean."

Source: Self-developed.

Propensity to Recommend

Also capitalizing on the depersonalization effect, this item is used to measure the propensity to recommend the town for its good services.

SAT5: "I would recommend St-Jean as a town where residents are highly satisfied with the municipal services provided."

Source: Based on item "Recommend XYZ to someone who seeks your advice" (Parasuraman et al., 1994a, p. 45).

Socio-Demographic Measures

According to Alreck and Little (1995), there are three purposes to collect information on the socio-economic and demographic profile of respondents. These purposes are (a) to "portray the nature of the sample", (b) to compare the sample to the population of interest and assess its representativeness, and (c) to provide with keys on which to segregate the data set into mutually exclusive clusters to assess group-specific influence (p. 168).

To compensate for an already long questionnaire the number of such items was restricted to seven and categorical or ordinal scales were used. Except for *homeownership status*, the included items are typical of socio-demographic items suggested by Alreck and Settle (1995).

Employment

EMPL: "What is your current **employment** status?"

Responses: 1. Employed full-time; 2. Employed part-time; 3. Seeking employment; 4. Retired; 5. Other.

Source: Self-developed item. Formed from the 1991 Census questions 30-34 & 44 to assess the issue of employment status (Statistics Canada, 1991a, p. 57). A question and reply structure free of conditional branching appeared more parsimonious than the source, and thus preferable.

Gender

SEX: "Gender."

Responses: m. Male; f. Female.

Source: Adapted from 1991 Census question 4 (Statistics Canada, 1991a, p. 45).
The use of the substantive *gender* was preferred to *sex*, which was used in the source.

Education

EDUC: "Please indicate the highest education level that you have completed."

Responses: 1. Elementary school; 2. High school; 3. CEGEP or equivalent;
4. University.

Source: Self-developed item. Formed from the 1991 Census questions 24-26 & 28 to assess the issue of completed education (Statistics Canada, 1991a, p. 55). A question and reply structure more parsimonious than the source, free of conditional branching, appeared sufficient.

Residence Mobility

STAB: "How long have you been living in Saint-Jean?"

Responses: 1. Less than 1 year; 2. 1 to 4 years; 3. 5 years or more.

Source: Self-developed item. Formed from the 1991 Census questions 20-22, (Statistics Canada, 1991a, p. 53). Intended to assess differences stemming putatively from the length of residence.

Homeownership

OWNE: "Do you or a member of this household own this dwelling (even if it is still being paid for) or do you rent it (even if no cash rent is paid)?"

Responses: 1. Own; 2. Rent.

Source: Adapted from 1991 Census question H2 (Statistics Canada, 1991a, p. 65).
Slightly reworded to blend in with the style of the other items.

Family Income

INCO: "Please indicate your total family income."

Responses: 1. Less than \$10,000; 2. \$10,000 to \$19,999; 3. \$20,000 to \$29,999; 4. \$30,000 to \$39,999; 5. \$40,000 to \$49,999; 6. \$50,000 to \$59,999; 7. \$60,000 or more.

Source: Based on question 45 of the 1991 Census (Statistics Canada, 1991a, p. 64). This item elicits a response on an ordinal scale with the expectation that it would appear less threatening to respondents and easier to complete.

Age

AGEG: "Please indicate your age group."

Responses: 1. 0 to 24 years; 2. 25 to 34 years; 3. 35 to 44 years; 4. 45 to 54 years; 5. 55 to 64 years; 6. 65 years or more.

Source: Self-developed item. Developed from question 3 of the 1991 Census (Statistics Canada, 1991a, p. 45). This item elicits a response on an ordinal scale with the expectation that it would appear less threatening to respondents and easier to complete.

Sequencing of the Items

The survey questionnaire presented the items in the following order: (a) quality items, (b) satisfaction items, and (c) socio-demographic items. A thank you note as well as an invitation to add comments to the replies, if desired, concluded the questionnaire. This sequencing of items responded to specific methodological imperatives.

Respondents may view the topics surveyed by some socio-demographic items, such as income, as threatening and may not reply to the item, or worse, as suggested by Davis and Cosenza (1993), may discontinue completing the survey. To minimize the impact of such non-responses, Alreck and Settle (1995) proposed that the "demographic items are ordinarily clustered together . . . and included at the end of the questionnaire" (p. 168). Thus, these items formed the last section in the survey questionnaire.

The positioning of the quality and satisfaction items raised the issue of the significance of the order in which items are presented. Conventional wisdom is that “respondent consistency may cause the ordering of the questions to affect the responses” (Scheaffer, Mendenhall, & Ott, 1996, p. 62). In a study on the satisfaction with municipal services, McClendon and O’Brien (1988) found that “the order effect occurred for only a small minority of respondents”, those for whom the issue at hand had only a low level of salience. However, in a more recent research, Schwartz and Hippler (1995) found that “respondents under mail conditions . . . were not significantly influenced by the order in which the questions were asked”, indeed positing “that respondents either read ahead or returned to previous questions once they encountered the subsequent one” (pp. 94, 96). Thus, although municipal services are prone to low levels of salience, suggesting that the quality/satisfaction question order is significant, conducting the survey by mail in fact distils the importance of ordering. Therefore, it was concluded that the order of quality and satisfaction items was not critical.

Published studies examining satisfaction or behavioural intention items in addition to the SERVQUAL were reviewed, seeking the solution to the order issue by applying the wisdom of seasoned researchers. Publications that specified the order of questions elicited the responses to quality items first (Brown et al., 1993; Teas, 1993). Consequently, quality items were brought to the attention of respondents prior to the satisfaction items.

Packaging

The survey package was composed of a transmittal letter, the questionnaire, and a self-addressed postage paid return envelope. The transmittal letter was in both English and French and was printed on Concordia University letterheads. The text of the transmittal letter is attached in Appendix B3.

Only the French version of the questionnaire was included in the package. Because more than 98% of the population in the survey area was French speaking, this arrangement was convenient to almost all potential respondents. To cater to the needs of English

speaking residents, arrangements were made with town officials whereby English questionnaires were available at City Hall, for exchange only. The transmittal letter indicated how the exchange could be arranged. Only one English questionnaire was exchanged; this questionnaire was completed and returned.

While no monetary incentives were included in the package, several attempts were made to induce replies by motivating respondents. First, the transmittal letter covered the “most likely questions that might arise in the recipient mind” (Alreck & Settle, 1995, p. 195). The appeal specified the “societal utility” (Davis & Cosenza, 1993, p. 284) of the research, to which the general public is generally responsive. The letter was signed by this researcher and by his thesis supervisor, because higher level signatory significantly improve the response rates (Labrecque, 1978). Both signature blocks included the persons’ names and titles in addition to the ink signature as recommended by Labrecque (1978). Second, business reply return postage was affixed to the return envelope so that respondents bore no costs. Third, as will be specified when discussing the collection procedure, follow-up notices were used.

Pre-Testing of the Instrument

This stage aimed at increasing the validity of the instrument by identifying and resolving wording and format problems in the instrument (Davis & Cosenza, 1993). As the questionnaire was translated into French by this researcher, this stage also aimed at ascertaining the consistency of the French with the English version.

The questionnaire and the covering letter were pre-tested on a convenience sample of 28 persons, comprising students, relatives, friends, and members of the faculty. It is believed that these persons were typical residents of towns, possessing a variety of attributes and a “frame of reference” (Davis & Cosenza, 1993, p. 204) similar to the survey sample. Feedback was appreciated and used to improve the instrument. Several members of the above mentioned group examined both linguistic versions of the instrument.

Data Sources and Collection

Because this research is concerned with the attitude of the residents towards quality of municipal services, a municipal setting was required. The desired municipal setting had to meet some qualifying conditions. These conditions were (a) the provision of a wide array of municipal services, so that all residents have an exposure to at least some of these; (b) a unique delivery agency, conspicuous to all residents, to avoid inter-jurisdiction confusion; (c) a homogeneous delivery of service, so that similar levels of service be available to most respondents; (d) a somewhat dense, urban-type population, to preclude distance being an obstacle to service experience; (e) linguistic concentration of the residents into a single language, to simplify the logistics of the administration; and (f) an interest from the municipal authorities. All the above conditions were met by the Town of Saint-Jean-sur-Richelieu, hereafter abbreviated to Saint-Jean, a town in Western Québec with a population of 32 000 residents.

The officials of the town were briefed *a priori* on the objectives and procedures of the research. No financial support was obtained from the city. The city's collaboration took the form of facilitating the exchange of French to English questionnaires and publicly supporting the project, by such means as attending the launch press conference. The survey results in aggregate form were made available for their perusal. This collaboration arrangement without commitment from the investigator is typical to conditions prevailing in other scholarly research with a similar focus (Goodsell, 1981).

Population

The population of interest for this study is the complete set of all elements, namely all residents of Saint-Jean. The population is operationally defined as the adults (18 years or over) living in households in Saint-Jean at the time of the survey. The respondent needed not qualify as a voter or demonstrate a minimum length of stay, as is sometimes required (Alreck & Settle, 1995).

Sampling Design

The survey was implemented through a mail delivered, self-administered questionnaire. This mode of administration was selected because mail surveys enjoy low data collection costs (Folz, 1996), optimizing the potential sample size for a given, limited budget. In addition, Alreck and Settle (1995) indicated that mail survey enable the survey to be conducted anonymously, with a lower “degree of psychological threats” (p. 37) facing the respondents. However, surveys conducted through mailed questionnaires often display low response rates (Scheaffer et al., 1996).

The sample units for this survey were residential households, including dwellings in apartment blocks. Using households as sampling units guaranteed that there were no overlapping, because they are physically distinct and no resident lives in more than one dwelling. Further, this sampling unit prevented redundancy, each different household containing different residents who hold their individual opinions. Sampling units could contain one or more elements. Some have argued that the specification of a household unit may be too broad, because “different family members may hold different opinions about the issue [at hand]” (Alreck & Settle, 1995, p. 55). While this assertion has merits, the current specification is a reasonable compromise, because it would be uneconomical to develop a sampling scheme that narrows the unit to the individual level.

In similar circumstances, some surveys call for the head of the household to complete the questionnaire. However, it is not felt that this person is more knowledgeable than other members of the household are with regards to many municipal services. In fact, it appears preferable to gather the opinion of random “adults so that the total respondents represented a proper proportion of each sex and age category” (Hatry et al., 1977). Thus, the covering letter specified that any adult in the household could respond to the survey, increasing the possibility that the replies be representative of the population.

The sampling frame is the collection of households in Saint-Jean, as available through un-addressed mail delivery to households. This frame is all-inclusive, because

almost all households fall in one of the mail carriers delivery routes, including the general delivery and rural routes. The only exceptions are (a) the homeless, (b) households that do not receive mail, and (c) any household for which all eligible members were temporarily absent for the entire survey period. These were presumed to be few, if any at all. In addition, the frame is exclusive, because no recipient on the household mail routes is not an household. All entries in this frame are households and no businesses are included.

Sampling Plan

One of the most popular sampling methods is the cluster sampling (Davis & Cosenza, 1993). The motivation behind this approach is mainly economical efficiency, because delivering un-addressed mail to sampling units that are in closely defined areas, such as mail carrier routes, is much more economical than using a frame where units are specified by address. Consequently, cluster sampling was used for this study.

The mail carrier routes were the cluster units. The clusters contained a mix of apartment blocks and houses. There were 29 mail carrier routes in Saint-Jean, ranging from 136 to 1018 households (Canada Post Corporation, 1996). The average number of households per cluster was roughly 600.

Using the clustering approach imply that all clusters look alike, but that much diversity is prevalent within the clusters (Scheaffer et al., 1996). However, 6 of the 29 clusters were considered by an expert to be alike, but distinct from the 23 other clusters. These clusters were located in the historic part of the town. According to the expert, they were populated mostly by less affluent senior citizens. To guarantee that these clusters were represented in the final sample according to their demographic weight, a stratified cluster approach was used (Davis & Cosenza, 1993). These clusters formed Stratum A, which contained 3858 of the 17396 households. Consequently, they received 22.2% of the instruments and stratification eliminated the risk that none, or all, or some inappropriate proportion of surveys be distributed to them by chance.

Thus, clusters were selected at random within each of the two strata, using a random number table, until the required sample size had been met. Two clusters were selected from Stratum A and three from Stratum B. To avoid exceeding the allocation for the stratum, the last cluster selected for each stratum was covered only partially, such that the number of surveys allocated to that cluster was less than the number of households in the cluster. A cursory examination of the geographic location of the selected clusters revealed that they were widely dispersed over the territory of the town.

Collection Procedure

The survey packages were delivered to households in the selected mail carrier routes as un-addressed mail. The delivery took place on a Saturday, in keeping with Canada Post policy concerning this type of mail. No package was returned intact and there is no evidence that any were left undelivered.

The distribution of the survey was accompanied by a media blitz to increase the awareness of the possible respondents. Prior to the distribution, a press conference was held. The same day as the press conference, this researcher appeared on a television show of a local community network to explain the research objectives. Print and electronic accounts of the press conference appeared around the delivery period.

Call-backs to non-respondents have been demonstrated to be effective in motivating some laggards to turn in their replies (Scheaffer et al., 1996). Because it was not possible to discriminate between respondents and non-respondents, the call-backs took a general form. A press release was issued in the second week after delivery, reminding about the importance to reply. It was published in some media.

Statistical Analysis

The data collected for this study was examined with four clusters of statistical procedures, including (a) descriptive statistics, (b) model estimation procedures, (c) tools

to assess the psychometric properties of the scales, and (d) tests of the research questions. The initial statistical treatment of the data was done with the SPSS software, version 6.1.1. for PowerMac (SPSS, 1994). Model estimation analyses were performed using the EQS Software, version 5.4 for PowerMac (Bentler & Wu, 1996).

Both quality batteries are measured through Likert scales. Likert scales are at the ordinal level of measurement (Davis & Cosenza, 1993), when narrowly defined. However, Davis and Cosenza (1993) suggested to relax the assumption of equal interval and to treat such scales "as though they are interval measurements" (p. 169). Nunnally (1978) stated that he "strongly believes that it is permissible to treat most of the measurement methods in psychology and other behavioural sciences as leading to interval scales" (p. 17). In addition, Likert scales were recommended to form summative measures (Alreck & Settle, 1995; Nunnally, 1978). The mean and variance of such composite variables can be computed and these two parametric procedures also apply to their components (Mulaik, 1972), implying that the scales are viewed as interval. In practice, SERVQUAL quality items were analyzed through an array of methods, such as linear regression and factor analysis, and never with nonparametric statistics. Thus, there is a consensus that these measures are considered as interval.

Descriptive Statistics

For the scales considered as interval as well as for the ratio-scaled items central tendency and dispersion estimates are provided. These statistics are (a) the number of valid cases, (b) the mean, (c) the standard deviation, (d) the skewness, and (e) the kurtosis. In view of their lower level of measurement, only the number of valid cases and frequencies are reported for ordinal variables—overall importance of the dimensions and the socio-demographic items. The scores for the discrepancy index are also reported, along with the same statistics as the components.

Because all these measures are fallible, highlighting and interpreting them extensively is not warranted. However, the studies on municipal services using

SERVQUAL examined in the literature review reported and discussed some univariate descriptive statistics. To the extent that these facts may be informative, the comparison between the results of previous studies and this research may shed some light on the issues at hand. Consequently, univariate statistics on the indicators will be exclusively analyzed when some suitable link can be made to a previous finding.

Model Estimation

The simultaneous model as well as other models were assessed by *CFA*. The use of *CFA* requires the elaboration of a theoretical underlying structure to the scale under investigation, enabling the specification of the relationships between the indicators and the constructs, as well as between the constructs, prior to the assessment of the degree of fit of the model to the data (Mueller, 1995). The conceptual network forming this model having been substantiated, the use of *CFA* is justified.

Estimation Algorithm

Given an acceptable sample size, the choice of the estimation procedure for *CFA* is largely conditioned by the tenability of the multivariate normality (MVN) assumption in the sample and, if the assumption does not hold, by the type of departure from normality exhibited. The more popular fitting functions, maximum likelihood (ML) (Lawley & Maxwell, 1963) and generalized least squares (GLS) are normal theory based estimators. Although departures from normality yield unbiased parameter estimates, it may “affect tests of statistical significance”, amongst other detrimental consequences (Bollen, 1989, p. 418).

In this study the measures are discrete and therefore not MVN. According to Bollen (1989), model assessment in presence of the discrete data problem involve estimating the latent underlying continuous indicators, computing their polychoric, or polyserial, or tetrachoric correlation, and estimating using a distribution-free estimator (ADF). Other possible remedial actions include treating the indicators as continuous and estimating with ML or GLS using robust adjustments (Bentler, 1995) or increasing the number of levels in the variables.

Cattell (1956) first suggested collapsing several variables into composite measures, sometimes called bundles or parcels. Doing so increases the number of scale points and might improve the shape of the distribution of the composite over that of the components. The fit of a model can be improved by using parcels instead of the measured indicators (e.g., Rahim & Magner, 1995). While, Gerbing and Anderson (1988) mentioned that composite measures may be meaningfully *only* when the scale is unidimensional, bundling is not recommended by Mulaik (1972), who stated that “composite scores will obscure any potential complexity that exists within the composites” (p. 69). Selecting between the remaining two methods—using a distribution-free estimator with latent indicators or using a standard procedure with robust adjustments on discrete indicators—is a function of (a) the sample size, (b) the model size, and (c) the number of scale points of the indicators.

The ADF estimator is known to unduly reject the true model in small samples and therefore is only useful for very large samples, such as those containing 5000 cases or more (Hu, Bentler, & Kano, 1992, p. 355). In addition, computational constraints arising from estimating with ADF restrict the number of model variables in practice to a maximum of 40 (Bentler & Dudgeon, 1996). In this study, the sample size is low to moderate and the models are large, precluding the use of the ADF procedure.

Bentler and Chou (1988) posited that indicator measures with no less than four levels might be treated as continuous. Then, it is appropriate to use a normal theory based estimator such as ML or GLS. In addition, the EQS software can compute the robust standard errors and an appropriately scaled variant of the T statistic. According to Bentler (1995), these robust statistics “perform better than uncorrected statistics where the normal distribution assumption is false and better than [a distribution-free method] in all but the largest samples” (pp. 48-49). In addition, the Satorra-Bentler (1988) scaled T statistic (\bar{T}) can be trusted “under a wide variety of distributional misspecification” (Bentler & Dudgeon, 1996, p. 585). In EQS 5.4, the GLS procedure has not yet been enhanced with robust statistics (Bentler, 1995). So, the enriched ML procedure will be used for analyses.

Model Identification

An underidentified model contains one or more parameters that are not uniquely determined (Schumacker & Lomax, 1996), yields an infinite number of solutions and is thus indeterminate. Schumacker and Lomax (1996, p. 101) suggested some preventive measures against underidentification. The procedures are (a) assigning a scale to each latent variable through the use of *reference indicators*, (b) avoiding reciprocal relationships between the factors, and (c) constructing a simple model with a minimum of parameters estimated.

Formal identification rules have been devised. First, a necessary but insufficient condition for model identification is the *order condition*, which imposes that the number of unique values of the input covariance matrix exceed the count of the free parameters being estimated, as demonstrated by the available degrees of freedom (Bollen, 1989; Schumacker & Lomax, 1996).

In addition, Bollen (1989) devised a comprehensive set of sufficient conditions to establish model identification. However, these rules are only valid for the assessment of models in which the measurement errors are not correlated. When these rules are to no avail, the sufficient condition of identification must be demonstrated empirically. This can be accomplished by two methods. The first method is Wald's (1950) rank test of local identification, which is performed by the EQS software (Schumacker & Lomax, 1996). The second method is the verification that the estimated information matrix is non-singular, because local identification is only possible if this matrix is invertible (Bollen, 1989). This last test is also catered by the EQS software, which uses this inverted matrix for the computations of some estimates (Bentler, 1995, p. 81). Bollen (1989) mentioned that additional tests need not be performed unless "doubt still remains about identification status" (p. 251).

Model Specification

The first step in model specification requires constructing and estimating a baseline model, defined by Bollen (1989) as “the simplest, most restrictive model” (p. 269) that can be logically formed linking the indicators to their theory-driven dimensions. According to Bollen (1989) this model is formed with one factor per observed variable, no error, and a free correlation matrix between the factors. The baseline model is used in computing incremental measures of model fit.

The *two-step* modelling approach will be used in this study. First proposed by James, Mulaik and Brett (1982), this approach involves assessing the measurement part of the model distinctly from the structural part. This line of investigation aims to prevent the “interpretational confounding” which is likely to occur if both parts of the model are estimated simultaneously (Anderson & Gerbing, 1988, p. 418). Three or more indicators per factors are required when this approach is used (Hayduk, 1996).

A measurement model is specified next, expressing the relationship between the indicators and the constructs as structural relationships, such as the following typical structural equation based on LISREL (Jöreskog & Sörbom, 1988) notation

$$x_i = \lambda_{i1}\xi_1 + \dots + \lambda_{ij}\xi_j + \delta_i$$

where x_i is indicator i , any observed variable, λ_{ij} is the loading of factor j on indicator i , with one, and only one $\lambda_{ij} \neq 0$ for a simple structure, ξ_j is factor j , δ_i is the error term associated with indicator i , typically uncorrelated amongst themselves and with the factors.

This measurement model “specifies the posited relations of the observed variables to the underlying constructs, with the constructs allowed to intercorrelate freely” (Anderson & Gerbing, 1988, p. 414). Two forms of the measurement model will be required, an eleven correlated factors first-order model and a correlated model comprising satisfaction

and a second-order factor sub-model for each battery. Second-order factor models constitute an appropriate manner of shaping measurement models (Schumacker & Lomax, 1996, p. 64). This measurement model will include higher order factors and specify their relationships with the first-order dimensions, linking these independent and dependent latent variables in a hierarchical structure. The top section of this model will display three correlated factors. The hypothesis tested at this level is that the correlation between the dimensions which is observed in the measurement model is attributed to a “higher-order common construct” (Etezadi-Amoli & Farhoomand, 1996, p. 692).

Estimated next, the model of interest represents the specification of the theory-driven relationships between the indicators and the factors, and amongst the factors. The model of interest is different from the measurement model in that there is no covariation between the dimensions and also in that some hypothesized causal relationships between the upper order constructs are specified.

Depending on the results of model assessment, further re-specifications may be desirable. The model assessment process continues iteratively until some optimal model is achieved. Over and above statistical assessment, model re-specifications must be theoretically sound.

The simultaneous model is graphically illustrated in Figure 3.1. From a methodological perspective, this model is a total disaggregation second-order factor model and it “provides the most detailed level of analysis” (Bagozzi & Heatherton, 1994, p. 42) of the information.

Using both the expectations and the perceptions items in the same model raises the issue of the independence of observations. Although the measures are obtained contemporaneously, in fact they represent repeated measures in a conceptually longitudinal model, where expectations and perceptions occur prior and after a treatment, that is during the service episode. According to Bentler and Chou (1988), “serial correlation among the responses are quite likely” in longitudinal models (p. 166). Other authors see correlated

errors of measurement as “an important feature” of measures repeated over time (Marsh & Grayson, 1994). Consequently, correlation between the error terms is built into the model to be estimated.

The loading of one indicator is fixed to the value of one for each of the 11 constructs. This procedure is carried out to resolve scale indeterminacy, such that the scale of each factor is set equal to that of the selected observed variable. Although the same result can be achieved by fixing instead the variance of the factors, Schumaker and Lomax (1996) indicated that the use of such “reference indicators” is preferable when the model contains indicators of both exogenous and endogenous latent variables (p. 100).

Assessment of the Factor Pattern

Standardized factor loadings are the parameter estimates resulting from model estimation that are retained for analysis. These statistics will be used for reliability and validity assessment. Bentler (1995) cautioned that interpretation of the standardized solution might be as problematic because “previously fixed parameters will take on new values . . . [, and] certain parametric constraints that were met in the problem may no longer be met in the solution” (p. 98).

Also of interest, a measure of the effectiveness of the model will be obtained by computing the variance explained by the model. For recursive models, the models’ squared multiple correlation coefficient associated with the equation is computed as one minus the square of the standardized parameter for the structural disturbance (Bentler, 1995). This statistic is sometimes designated as the Coefficient of Determination and identified as R^2 . Rigdon (1995) defined recursive structural models as those which do not display “a reciprocal relationship, a feedback loop, or correlated disturbance terms” (p. 361). Incidentally, all structural models used in this study are recursive.

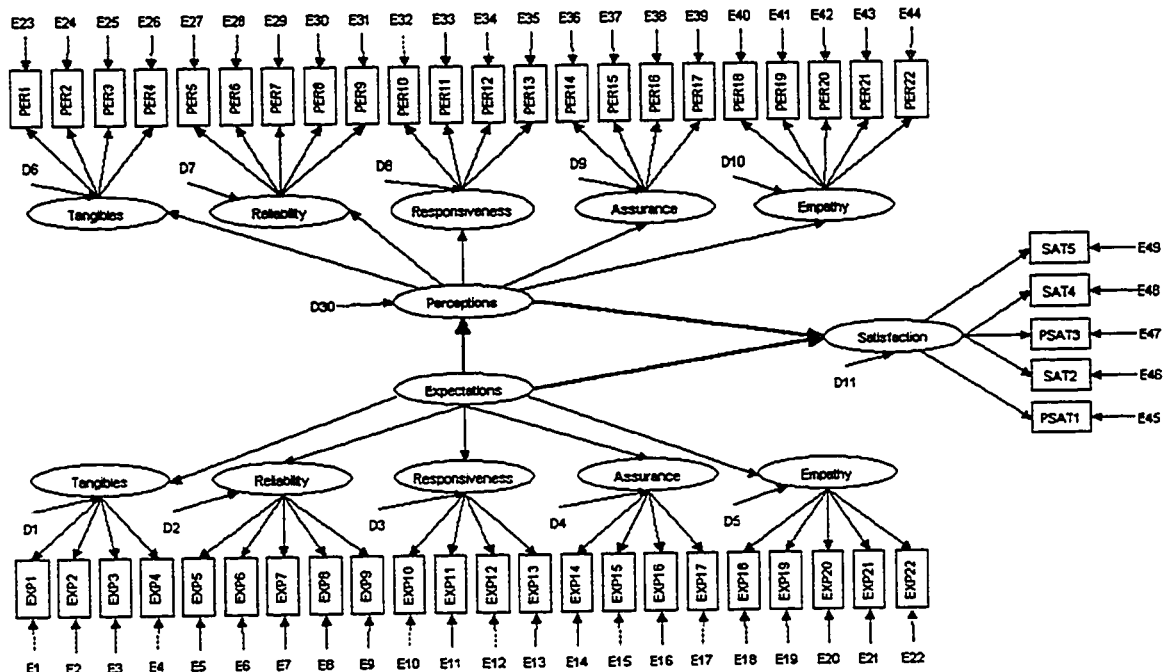


Figure 3.1. Illustration of the simultaneous model. Correlated errors not displayed.

Assessment of the Model Fit

A subsequent level of assessment involves a judgement on each model's fit. There is a selection of fit indices that have been used for this purpose and some are routinely printed on the EQS output. Many indices are normed to vary from zero to one, with one representing a perfect fit and a .90 score "often . . . considered to be indicative of a good overall fit" (Mueller, 1995, p. 90). However, the different indices measure slightly different things. Various indices differ in their components and how they relate them, yielding different and sometimes contradictory results. An improvement on one index, when comparing a model to another, may not be accompanied by a change of the same magnitude on another index, which may possibly deteriorate. Hence, comparing models by their indices is not an exact science. Bollen and Long (1993) posited that it would be imprudent to judge a model on a single index. Moreover, the model fit is better assessed when several measures are examined in perspective (Marsh, 1995).

According to Schumaker and Lomax (1996), model assessment indices fall into three categories: (a) model fit criteria, (b) model comparison criteria, and (c) model parsimony criteria. The model fit criteria that will be reported for this study are (a) the \bar{T} statistic, (b) the goodness-of-fit index (GFI), and (c) the robust root-mean-square error of approximation (RMSEA*). The model comparison criterion will be (d) the robust comparative fit index (CFI*). The model parsimony criteria will be (e) the robust normed chi-square (NC*) and (f) the robust parsimonious CFI (PCFI*). These indices form a representative assortment of the available measures of fit. Robust indices are designated by their acronym followed by an asterisk in keeping with the notation used by Byrne (1994) for the robust CFI, "CFI*" (p. 95). For \bar{T} , however, *-bar* is used instead of an asterisk to fall in line with the manner this index is generally designated.

The probability that the model's implied covariance matrix equals the sample's covariance matrix is assessed by finding the *p-value* of the \bar{T} statistic, which is distributed as a χ^2 , using the model's degrees of freedom. Consequently, the \bar{T} index will be reported

along with its associated degrees of freedom and its probability, from a χ^2 distribution table. This index is robust to departures from MVN (Bentler, 1995). A non-significant probability, along the “conventional” *p-value* > .05, is interpreted as a good model fit (Bollen, 1989, p. 266). However, Schumaker and Lomax (1996) mentioned that the *T* index “is sensitive to sample size, because as sample size increases (generally above 200), the . . . test has a tendency to [unduly] indicate a significant probability level” (p. 125). This propriety is shared by the \bar{T} index. Because the sample size for this study is in excess of 200, the probability associated with this index is likely to be biased. In addition, this index is detrimental to “models with many variables and degrees of freedom [that] will almost always have significant chi-squares” (Rahim & Magner, 1995, p. 125). Consequently, this statistic is reported for its use in the computation of other statistics.

The GFI (Jöreskog & Sörbom, 1988) is reported in view of its wide popularity. This is a normed index. This statistic is not corrected for departures from normality. It must thus be considered with circumspection.

The RMSEA* statistic is not read from the output, but computed using the formula developed by Steiger and Lind (1980, as cited in Marsh and Balla, 1994) and re-expressed by Rigdon (1996, p. 374) as

$$RMSEA = \sqrt{\frac{T - df}{(N - 1) \times df}}$$

and substituting in \bar{T} instead of *T*. Browne and Cudeck (1993) subjectively assessed values of RMSEA around or below .08 to indicate a “reasonable” (p. 140) error of approximation. However, a more stringent upper limit of .05 is more common (Raykov & Widaman, 1995; Schumacker & Lomax, 1996). This index is not influenced by the sample size (Marsh & Balla, 1994).

Establishing the CFI* value requires computing the index developed by Bentler (1990, p. 241) and re-expressed by Rigdon (1996, p. 374) as

$$CFI = 1 - \frac{T_k - df_k}{T_b - df_b}$$

where the numerator is the T statistic minus the number of degrees of freedom of the theory driven model of interest and the denominator is the parallel values for a restricted baseline model. This model implies the restrictions that the denominator is larger than the numerator and that both are greater than zero. Again, a robust estimate is computed by substituting \bar{T} to T (Byrne, 1994, p. 95). Because this is also a normed index, the value of .90 constitutes an acceptable threshold (Byrne, 1994, p. 55), but .95 has also been suggested to “reduce the number of severely mis-specified models that achieve acceptable values on this criterion” (Rigdon, 1996, p. 374). The CFI statistic was empirically verified to produce consistent estimates even for small samples when analyzing non-normal data (Wang, Fan, & Wilson, 1996, p. 243).

The NC* index is computed by dividing the \bar{T} statistic by the number of degrees of freedom (Schumacker & Lomax, 1996). There are diverging opinions as to what constitutes a rule-of-thumb acceptable value for this index. Values as large as five have been seen as indicators of acceptable fit (Mueller, 1995, p. 84), but a maximum value of two is more often suggested, while not justified (Marsh, Balla, & McDonald, 1988). The very conservative value of one is also mentioned (Ding, Velicer, & Harlow, 1995).

The PCFI* value is obtained by applying the parsimony adjustment for normed indices (Mulaik et al., 1989) to CFI*. The parsimony index is computed as the division of the number of degrees of freedom of the theory driven model of interest by the number of degrees of freedom of the restricted baseline model. Rigdon (1996) suggested a score of .75 or above as indicative of parsimony.

Nested Models Assessment Procedure

A further assessment tool is available to determine exactly if incremental model re-specifications are worthwhile, provided that the models are nested. One model is nested within another model when one or more free parameters of the freer model are fixed in the restricted model. For the comparison of nested models, the likelihood ratio test can be performed (Bollen, 1989). This is achieved by subtracting the value of the \bar{T} index for the freer model from the parallel value from the restricted model. This value is distributed as a χ^2 , with the number of degrees of freedom equal to the number of fixed parameters. A significant score is interpreted as evidence against the restricted model (Bollen, 1989), implying that the model with more free parameters was a better fitting model. Freeing parameters instead of fixing them can be the basis of a similar test. In this case, a significant decrease in the value of the \bar{T} statistic is viewed as a model improvement (Byrne, 1994, p. 64).

The structural model is nested within the measurement model. Rindskopf and Rose (1988) asserted that a second-order factor model is nested within its correlated first-order factor model and that incremental fit between these two models can be assessed by the likelihood ratio test. Marsh (1987) also computed this difference. However, Byrne (1994) described the relationship between the models as "a special case" (p. 119) but conspicuously did not perform a likelihood ratio test. Bollen (1989) did not use this test either when assessing the fit of a second-order factor model. So, there appears to be diverging opinions between researchers on the appropriateness of using the likelihood ratio test in these circumstances, and fundamentally, on whether or not such models are nested.

Psychometric Properties of the Scales

The five expectations dimensions, the five perceptions dimensions and the satisfaction scale constitute eleven distinct scales whose psychometric properties must be investigated. The assessment of the psychometric properties of the scales requires examining (a) their factor pattern, (b) their reliability, and (c) their validity.

The factor pattern will be examined first, as some of its results are useful for the assessment of the reliability and the validity of the scales. Notably, it enables the investigation of the relationship between constructs “without the bias that measurement error introduces” (Steenkamp & van Tripp, 1991, p. 284). The validity and reliability issues are then covered, in order.

Factor Pattern

The factor pattern of the individual scales was examined through *CFA*. The procedure for this step used the same elements as those examined during the initial model estimation. The standardized factor loading estimates and a selected set of fit indices are reported to enable further validity assessment.

Parameter significance can be further assessed through two indices routinely provided by the EQS software. First, the Wald test index that suggests parameters whose setting to zero would improve the model fit. Secondly, the Lagrangian Multiplier (LM) index that suggests parameters to free in order to improve model fit. These indices are empirically driven and their results must be subordinated to conceptual soundness.

Validity

The validity of the scale is concerned with the absence of bias—non-random error—in a scale, and it is “multifaceted in that there are different kinds of validity” (Davis & Cosenza, 1993, p. 171). These different forms of validity are (a) content validity, (b) criterion-related validity, and (c) construct validity.

Content validity concerns the ability of the indicators to fully encompass the domain of the concept. The exploration required to specify the full domain of a concept may include a thorough literature review of the topic (Carmines & Zeller, 1979), as well as the solicitation of expert advice (Bollen, 1989). This form of validity is assessed qualitatively. Embedded is trait validity (Nunnally & Bernstein, 1994), a layperson’s assessment of the pertinence of the items to the concepts they render operational.

Criterion-related validity is concerned with “the extent of the correspondence” (Carmines & Zeller, 1979, p. 18) between the scale and an external variable, the criterion. This facet of validity is achieved through an assessment of the validity coefficient, computed as “the absolute value of the product-moment correlation between the test and a specified criterion” (Lord & Novick, 1968, p. 261). Bollen (1989) posited that this index may be an inappropriate base to assess validity. He stated that the index is influenced by the error in both the scale and the criterion and that different criteria entail different values of the index, “leaving uncertainty as to which is an accurate reading of a measure’s validity” (p. 188). Taking a slightly different perspective, Vogt (1993) defined criterion-related validity “as the ability of a test to make accurate predictions” (p. 52). On the issue of the predictive value of the criterion, Nunnally (1967) stated that selecting items on the basis of their correlation with a criterion is incorrect and that “a far better approach is to form a battery of tests from homogeneous measures of known factor composition” (p. 245). Thus, the utility of this form of validity as an empirical assessment tool appears useless, or at least dubious.

Construct validity is the extent “to which a construct achieves empirical and theoretical meaning” (Steenkamp & van Tripp, 1991, p. 283). Lord and Novick (1968) posited that “for scientific purposes, the most important characteristic of a test is its construct validity” (p. 278). The unidimensionality of a scale, as well as its reliability, are prerequisites to the assessment of construct validity (Ahire, Golhar, & Waller, 1996), that is further made operational through the assessment of convergent and discriminant validity.

A scale is unidimensional when indicators load exclusively on a single underlying factor (Gerbing & Anderson, 1988). The unidimensionality of the single factor scales is assessed through a flurry of unifactorial tests, with a measurement model being specified and estimated for each construct, and achieving a score of .90 or more (Ahire et al., 1996) on selected GOF indices. Anderson and Gerbing (1988, p. 415) stated that the presence of correlated errors of measurements in a model is evidence of a violation of the

unidimensionality of the factors. This means that within-dimension correlated errors preclude unidimensionality, but not the lagged correlation of error terms across successive waves of an instrument, which is typical in longitudinal designs. Reliability of the scale, the second necessary but not sufficient condition to construct validity, is examined distinctly *infra*.

Convergent validity examines the strength of the association between measures of a construct. Convergent validity rests on the validity of the indicators, which Bollen (1989) defined as “the magnitude of the direct structural relation between” (p. 197) the latent variable and the indicator. To assess the validity of an indicator, he proposed several methods, one of which is the standardized validity coefficient (SVC), λ^S , computed as the standardized loadings of the factors on their theoretical indicators (Bollen, 1989). Statistical significance of the loading coefficients is the weak condition to convergent validity, while large parameter estimates are a stronger condition (Steenkamp & van Tripp, 1991). In addition, a large value is required particularly from the CFI* index, that assesses the model of interest against a null model—a model with as many constructs as indicators.

Discriminant validity is convergent validity’s “mirror image” (Vogt, 1993, p. 71). Low association between the constructs of the scale traditionally demonstrate discriminant validity, with no correlation between factors as face evidence. When it is not the case, and more so as the correlation between two factors becomes large, it raises the issue of determining whether two, or only one, constructs are present.

A formal test was proposed by Anderson and Gerbing (1988) and empirically illustrated by Ware, Galassi and Dew (1990), who compared a two-factor correlated, a two-factor orthogonal and a single factor model formed with the same data. The orthogonal and the single factor models are nested within the oblique—correlated—model. Hence, likelihood ratio tests for difference of T statistic were performed, with the results distributed as a χ^2 with a single degree of freedom. A non-significant—large— T statistic difference evidenced the superiority of the oblique over the orthogonal or the single factor

model. For discriminant validity purposes, only the correlated to the single factor models need to be compared.

Discriminant validity for a scale comprising more than two factors is not tested simultaneously, but in pairs of factors (Anderson & Gerbing, 1988). Consequently, testing a scale using this method implies t iterations of the test (Ahire et al., 1996, p. 40), such that

$$t = \frac{n!}{(n-2)! \times 2!}$$

where n is the number of scales being tested, and t is the number of tests to perform.

To achieve a pre-set overall significance level, an adjustment is required for the significance level of each test within the family of tests. Anderson and Gerbing (1988, p. 416) specified the individual significance level to be computed as

$$\alpha_0 = 1 - (1 - \alpha_i)^t$$

where α_0 is the overall significance level, set deliberately at .05, because this level represents a typical desirability level, α_i is the significance level used for each test, and t is the number of tests to perform.

Reliability

The reliability of a scale is the degree to which it is free from random error, such that a “measuring procedure yields the same results on repeated trials” (Carmines & Zeller, 1979, p. 11). Reliability can be estimated in many fashions, such as test-retest, split-half, internal consistency, etc. For practical reasons, only the internal consistency method was used for this research.

Cronbach's α coefficient (Cronbach, 1951) was computed for all directly measured scales. For basic research, a scale achieving a reliability score around .80 is deemed reliable, while scores as low as .50 are acceptable for exploratory studies (Nunnally, 1967), such as this one.

When required, the individual item reliability is computed as the squared multiple correlation (SMC) for x_i —the square of the loading of the factor on the indicator (Bollen, 1989, p. 221). It should be noted that the SMC is the square of the previously mentioned SVC. In addition, it is also a lower bound estimate that discounts the contribution of the variance which is specific to the item. The SMC is computed as the square of the standardized factor loading estimate (Bollen, 1989).

Investigation of the Research Questions

In this sub-section, the tests to elucidate the research questions will be examined. The specifications of the research questions are summarized in Table 3.1.

Research Question 1

Do the data support a second-order factor quality model to represent residents' perceptions of quality of municipal services?

This research question will be partly answered by examining the model fit and the psychometric properties of the quality scales. It will be further examined through *EFA* and alternate factor pattern estimations.

Although *CFA* is preferred to “the conventional *EFA* due to its conceptual strengths” (Ahire et al., 1996), an *EFA* can be performed to examine the patterns in the data when estimated without the constraint of a model, seeking some congruence of the factor patterns when estimated through these two approaches and to compare the findings of this research to previous studies. *EFA* was performed with the SPSS software. The ML extraction algorithm was selected for congruence with the *CFA*.

The appropriate number of factors to retain varies according to the selection criteria utilized. The more popular criteria are (a) Kaiser-Guttman's arbitrary rule of eigenvalues with a score equal or in excess of unity; (b) the interpretation of the scree plot, which was suggested by Cattell (1966); (c) the cumulative percentage of variance extracted exceeding some threshold, often set at 60% for social sciences (Hair, Anderson, Tatham, & Black, 1992); and (d) the *a priori* specification of the number of factors on some theoretical basis, a deterministic approach that contrasts with the other three, which are data emergent. When two or more factors are extracted, rotation enhances the interpretability of the coefficients' matrix (Norusis, 1988). The OBLIMIN rotation is used in all applicable cases because the theory calls for correlated factors.

Model estimations with different numbers of factors will be examined. In addition, a CFA model with a single factor per quality battery will be assessed.

Research Question 2

Do the data support equal or different relative importance levels for the quality dimensions?

This research question will be addressed by examining the weights of the dimensions in CFA. The focus will be two-fold. First, the equality or inequality of the weights of the factors within a battery; second, the equality or inequality for parallel factors across the batteries.

On the first issue, according to Hair et al. (1992), the CFA standardized solution's parameter estimates are similar to regression's standardized beta coefficients, "allowing for a direct comparison between coefficients as to their relative explanatory power of the dependent variable" (p. 20). Free of multicollinearity because the structural model imposes orthogonal relationships between the dimensions, the standardized factor loading parameter estimates can be interpreted as the relative weights of the dimensions in shaping the second-order construct.

The relative importance of the dimension in the shaping of the higher order construct is thus a function of the magnitude of the parameter estimate, with a zero value indicative of no substantive effect and “increase in values correspond[ing] to increased importance in the causal relationships” (Hair et al., 1992, p. 450). Bollen (1989, p. 125) cautioned that the adequacy of standardized coefficients to importance measures is not automatic and constitutes a researcher decision.

The bulk of the evidence has been in favour of unequal weights of the dimensions in shaping the quality construct. This specific issue investigates if there is a significant difference between the dimensions weights within a quality battery. This will be determined by computing a likelihood ratio test for constrained equal and free models.

On the second issue, because the expectations and perceptions are measured distinctly, two sets of relative importance of the dimensions can be computed. Thus, the simultaneous model has a somewhat longitudinal profile for quality batteries, with expectations representing measures prior to the delivery and perceptions being measured after. Bollen (1989) hypothesized equivalent measures for each parallel component of a longitudinal study and “this substantive idea is translated into the model” (p. 234) by imposing constraints on the appropriate parameters. Imposing such constraints would imply that respondents form a unique image of quality for both phases of the assessment. Marsh and Grayson (1994) stated that factorial invariance over time may be safely assumed when the time intervals are short and outside of “a developmental period in which change is expected” (p. 334). These two conditions are met in this study. In addition, these constraints provide the advantage of reducing the number of free parameters to estimate, thereby reducing the likelihood of computational problems. Hence, these constraints seem appropriate both from a theoretical and a methodological perspective.

However, residents may nonetheless form an attitude towards expected quality differently than towards perceived quality. In that case, the structural as well as the measurement parameter estimates may not be analogous in both second-order factors.

Consequently, the empirical appropriateness of parallel constraints will be assessed via a likelihood ratio test. Should this test fail, this will be interpreted to mean that residents shape expected quality differently from perceived quality.

Research Question 3

Do the data support the satisfaction model to represent residents' satisfaction with municipal services?

This research question will be answered by examining the model fit as well as the psychometric properties of the satisfaction scale.

Research Question 4

Do the data support the simultaneous model or the perception-only model to represent the relationship between quality and satisfaction with municipal services?

This research question will be partially answered by examining the model fit and the psychometric properties of the simultaneous model and the perception-only model, as well as incremental fit measures. Subsequently, the data will be partitioned into homogeneous sub-sets along selected socio-demographic attributes of the sample. Multi-group comparisons will be performed to elucidate possible discrepancies motivated by such respondents characteristics.

Table 3.1.
Summary of the Research Questions

Research Question	Issue	Statistical Test	Implications
1 Quality	Fit of the second-order factor model	Assessment of model fit and psychometric properties Exploratory factor analysis Specific model tests	Dimensionality of SERVQUAL Universality of SERVQUAL
2 Relative importance of the quality dimensions	Equal or unequal weights within a battery Equal or unequal weights across batteries	Likelihood ratio tests	Rank-ordering of dimensions Single or dual set of importance weights
3 Satisfaction	Fit of the first-order factor model	Assessment of model fit and psychometric properties	Satisfaction scale suitable for local services
4 Quality influences satisfaction	Simultaneous model or perception-only model	Assessment of model fit and psychometric properties Multigroup likelihood ratio tests	Causal link between the attitudes Pertinence of measuring expectations Viable alternative to gap scores

RESULTS AND DISCUSSION

In this section, the outcome of the survey will be examined. First, the steps that were taken to ensure the exactness of the data set are reviewed. Then the representativeness of the sample is ascertained, followed by univariate statistics of the data set. Next, the actions required in the making of the data suitable for multivariate analysis will be detailed and the model of interest will be estimated. Finally, the psychometric properties of the scales will be submitted to testing and the research questions will be considered.

Coding and Validation of the Data

In keeping with Alreck and Settle (1995), replies were numbered sequentially upon arrival and the postmark date was recorded for each case. The postmark date was considered the best available proxy to the date of mailing.

Some respondents attached comments to their replies or wrote them directly on the questionnaire. These comments were recorded *verbatim*, but all nominative information was removed. Because the qualitative assessment of this information is outside of the scope of this study, none was performed and these comments are not reported. All other enclosures were discarded.

Data Entry

A double data entry was performed by a specialized third party, according to the file specifications and the detailed instructions provided. If available, the single selected answer to each item was entered in the appropriate field. All fields were completed—0s were added to the left of a number if required to fill a numeric character field of fixed length. Missing answers were left blank and later coded as 9s, or 999s for three numeric characters fields. Any other prevailing condition prompted an anomalous code entry, an 8 or an 888.

Data Accuracy

The accuracy of the data set was evaluated in three steps: (a) verification of the completeness of the data set, (b) audit of a sample, and (c) assessment of the likelihood of the descriptive statistics, as well as the audit of anomalous and missing codes.

Alreck and Settle (1995) suggested that the raw data file be first examined to determine if all cases were present and complete. A visual inspection revealed that the data set formed a rectangular matrix, with the proper number of rows and columns, and that all positions within the matrix were filled.

The contractor guaranteed that the keying error rate is less than one half of 1%. An audit was performed to assess this claim. A one-in-ten systematic sample of the raw data file was sight verified against the questionnaires for accuracy. Case numbers ending with a 5 were selected for examination through a random procedure. No errors were found. Thus, there is at least a 99% probability that the error rate $\leq .005$.

Tabachnick and Fidell (1989) suggested that additional tests of data accuracy be performed. These tests are (a) inspecting the range of all the variables in the file for out-of-range numbers, (b) ascertaining the plausibility of measures of central tendency and of dispersion, and (c) proofing the use of anomalous and missing items codes. The first two tests were conducted through a visual inspection of the frequency distributions and other descriptive statistics of the variables. Anomalous and missing data codes were 100% proofed against the questionnaires. No discrepancies were found on any of the three accounts. Consequently, the electronic data file was deemed to be accurate.

Anomalous Data

Anomalous data result from mistakes made by the respondents in completing the questionnaire, such as providing two different answers for the same item. Such anomalies are mainly attributed to distraction or negligence on the respondent's part. The types of anomalous data that were observed in the data set were (a) multiple coding, (b) unmarked responses, and (c) inadmissible replies. They are discussed below.

Multiple Coding

Multiple coding occurs when respondents provide more than one answer to a specific item or an answer that does not conform to the answer format, such as striking several numbers when the expected answer was a circle around one number only. The contractor entered the anomalous value code as the multiple value code. In all, there were 43 occurrences of multiple coding in the raw data matrix, which constitutes less than 0.17% of the fields in the data file. For multiple replies, the following resolution rules were applied:

1. For sections where the replies were Likert scaled—expectations (EXP1 to EXP22), perceptions (PER1 to PER22), and satisfaction (SAT1 to SAT5)—the rounded average of the multiple replies were recorded (19 occurrences);

2. For overall importance (IMPMOS, IMPSEC, and IMPLEA) and socio-demographic (EMPL, SEX, EDUC, STAB, OWNE, INCO, and AGEG) sections, the corrected value is missing, because the exact answer could not be determined nor estimated (19 occurrences).

3. For one occurrence, these rules were disregarded because the item with the double reply was followed by a missing reply item. This occurred in the satisfaction section, where one item was negatively worded while the other was positively worded. Due to the close proximity of the two problematic items, it was safe to assume that one of the two selections for the first item was the reply to the next item. Which value to assign to which item was resolved by examining the pattern of replies to the other questions in this section. This judgmental approach was deemed superior to averaging for one item and later imputing a value to the other item, because both values would then be different from the replies. The current method safeguards the actual replies and assigns them to the most likely item.

4. Unexpected forms of answer, such as tic marks, were assimilated to multiple answers, but could not be assessed with rules. Therefore, each case was examined on its merits. In two occurrences, the intended answer was identified and recorded. For the other two occurrences, a missing value code was recorded.

Unmarked Responses

For the relative importance items (IMPTAN, IMPREL, IMPRES, IMPASS, and IMPEMP) battery, not all respondents had written a zero to indicate that they attributed no importance to a particular dimension. Consequently, these occurrences were unduly recorded as missing. The value zero was substituted in, except when replies were missing for all five items, which were left unchanged at this time.

Inadmissible Replies

The replies to the overall importance items stipulate a rank order preference. Hence, they need to be mutually exclusive. One case had the same reply for all three items and another case for two items. Because it was impossible to discriminate which single item deserved the provided reply, a missing value was attributed for the five occurrences.

Replies to the relative importance items and the overall importance items were consistent at the aggregate level. However, a detailed examination revealed some individual inconsistencies between the levels of the values assigned to the dimensions in the first set and the rank ordering in the second set. Because the only available remedial measure was to create more missing by substituting in the missing code to delinquent values and because the overall importance items had no analytical use otherwise, no action was taken.

Scanning the replies to the relative importance items revealed that 45 cases did not total 100 on the constant-sum scale. This deficiency was attributed to the higher level of difficulty implied in providing a reply to a constant-sum scaled item. Because this anomaly afflicted 12% of the cases, assigning missing values was not a viable option. Proportioned values were computed and rounded to fit the constant-sum scale 100 total rule.

Once all these remedial measures were taken to solve anomalous data, the data set was suitable for analysis although it contained some missing values.

Sampling Results

A total of 391 questionnaires were returned for a response rate of 20%. In past similar studies, response rates for residents' surveys have reached the level of 37% (Carlson & Schwarz, 1995) and 51% (Das et al., 1995). However, in general much lower response rates—5 to 10%—are observed for mail surveys, such that “response rates over 30% are rare (Alreck & Settle, 1995, p. 35). Furthermore, a recently reported study using SERVQUAL in the commercial sector reported rates on multiple samples ranging from 19% to 28% (Mels, Boshoff, & Nel, 1997). In this context, the 20% rate obtained in this study is not considered atypical.

Armstrong and Overton (1977) recommended procedures to quantitatively assess the representativeness of the sample. One way is to compare the sample to its population according to their attributes, notably their socio-demographic characteristics. Another one is to examine if tardy replies are similar to early ones. Both of these tests were performed and their results are reported in the subsequent sub-sections.

Comparison of the Sample to the Population

The socio-demographic attributes of the respondents are profiled in Appendix C. The results of the seven socio-demographic items collected from the sample were compared with the relevant attributes of the population, as inferred from the 1991 Census data (Statistics Canada, 1991b). A test for the likelihood of the sample belonging to the population was computed on each variable. To enable testing proportions, ordinal scales were converted to display dichotomous results. The number of cases for each test varies due to missingness. The results were summarized via the *z-value* scores and the probabilities of occurrence. Table 4.1. contains the referent data and the results of the tests.

Table 4.1.

Comparison of the Sample's and the Population's Socio-Demographic Attributes

Measure and group	Sample	Town	z-value
Employment status			
In the labour force ^a	74%	65%	3.7 ^{††}
Not in the labour force	26%	35%	
Gender			
Male	61%	49%	4.7 ^{††}
Female	39%	51%	
Education Level			
Elementary and high school	42%	59%	-6.81 ^{††}
CEGEP and university	58%	41%	
Stability of residence			
0-4 years	13%	25%	-5.43 ^{††}
5 years or more	87%	75%	
Ownership			
Own	78%	39%	15.13 ^{††}
Rent	22%	61%	
Family income			
0 to \$39 999 ^b	49%	52%	-1.32
\$40 000 or more	51%	48%	
Age group			
18 to 54 years ^c	71%	71%	-0.13
55 years or more	29%	29%	

^aThe "in the labour force" class includes the survey replies (1) employed full-time, (2) employed part-time, and (3) seeking employment. The "not in the labour force" class comprises (4) retired and (5) other. This classification is compatible with the 1991 Census status (STATSCAN, 1991a, p. 82).

^bThe \$40 000 dichotomization threshold was suggested by the population's central tendency measures: mean \$42 283, median \$38 535 (STATSCAN, 1991b, p. 859).

^cThe population size for ages 18 and 19 was interpolated from the Census data (STATSCAN, 1991b, p. 465).

^{††} $p < .01$, two-tailed.

The results for the age groups and the income levels variables show no significant difference between the sample's and the population parameters ($p\text{-value} < .05$), while other variables do. For gender, the results are significant, but opposite to what is usually observed. Bentler and Chou (1988, p. 168) cautioned that women are "more available and/or cooperative as research subjects", leading to their overrepresentation in the samples. In this instance the women are underrepresented, perhaps due to a persisting societal belief that men handle important household matters. When gender is considered along with age, the results do not coalesce either with Verba (1996), who stated that "respondents overrepresent the elderly and women" in opinion pools (p. 3).

The differences manifested for owners as compared to tenants, as well as for long-time residents as compared to those newly arrived, could be attributed to a higher level of civic conscience as well as a higher level of interest for municipal services' quality in the overrepresented classes. It could also indicate a greater opportunity for homeowners and long-time residents to have experienced and formed an opinion on services, making it more conducive for them to reply to the questionnaire.

Respondents with higher levels of education being overrepresented may indicate that they found it easier to answer the items than the less educated. Finally, persons in the labour force being overrepresented runs contrary to the expected results, because those not in the labour force certainly had more disposable time to complete the questionnaire.

In summary, the comparison of the sample profile to the population's reveals that their shape differ on some attributes, but that they are identical on the critical age and income levels determinants. However, this discrepancy is not *per se* evidence that the data lacks external validity. Further investigation for possible biasedness will be examined in the subsequent sub-section.

Early and Late Responses

The sample could be biased in favour of dissatisfied residents, because “dissatisfied customers may be more likely to respond, and respond negatively, to a satisfaction survey because dissatisfaction itself is inherently more action-oriented and emotionally intense than satisfaction” (Peterson & Wilson, 1992, p. 64). This threat can be investigated by estimating the replies of non-respondents. According to Armstrong and Overton (1977), if the responses of respondents and non-respondents to an item differ, tardy replies may be more representative of non-respondents than of early respondents. These authors based their statement on the assumption that persons who have an interest for the subject of the questionnaire tend to reply more quickly. To test for the presence of this non-response bias, they proposed to ferret the presence of trends in the replies by comparing aggregate results for successive waves of replies and examining time trends (Armstrong & Overton, 1977).

Comparison of Waves

The hypothesis that early and late responses differ can first be tested by extrapolation. According to Armstrong and Overton (1977), this procedure aims at examining the consistency in the replies over successive waves of replies. To assess this phenomenon, the data was dichotomized on the basis of replies received before or after the publication of the reminder. The first wave, WAVE1, consisted of the 300 replies postmarked before or on the day the reminder was published; the second wave, WAVE2, consisted of the 91 responses postmarked after that date.

A parametric test for difference of means was computed for all items in the expectations, relative importance, perceptions, and satisfaction sections. Only the PER19 item revealed to have a difference significant at the $p < .05$ level, with tardy respondents having a lower perception towards the convenience of operating hours. This evidence runs against the proposition that dissatisfied customers are more prone to answer.

In view of the categorical or ordinal scale measurements of the socio-demographic items, a non-parametric test was deemed more appropriate to assess for differences in the waves. Significant differences for gender and ownership were observed using the Mann-Whitney-Wilcoxon Rank sum test, with tardy replies containing significantly more responses by women and tenants than the early replies.

Several possible explanations come to mind as to why proportionately more women and tenants are represented in the late replies. First, the early replies may have come proportionally more from households where males hold the symbolic position of *head of household*. This hypothesis would also fit the sample's ownership variable because, for older couples, the man is often the designated owner for the household's real property.

Secondly, men and owners of real estate may have felt more concerned with the survey, or may have had more pre-eminent opinions to offer, and hence were overly represented in the early replies. This would fall in line with the proposition by Stipak (1979a) that some citizens may be prompted to give opinions about public services even if they have less knowledge about the services. In this case, the reminder was the prompting phenomenon for those who had resisted the temptation originally, but finally put in a tardy reply. The replies themselves, however, dispute the proposition that tardy respondents are less informed about the services, because there were no significant differences between early and tardy replies for all but one item.

Third, and last, women and tenants may be more prone to read the paper in which the reminder was published. Or else, they may have been more influenced by the appeal to reply contained in the message.

Trend Analysis

Armstrong and Overton (1977) also suggested the time trend procedure, which assesses the evolution of the replies as time elapses during the collection period of the single survey. They claimed that this procedure is superior to the previous one, because "the possibility of a bias being introduced by the stimulus itself can be eliminated" (p. 397).

A regression of the individual items scores on a deviated form of the return date (Babakus & Boller, 1992) set this test in operation.

Using the postmark information, the cycle time between distribution of the questionnaires and reception of the replies was measured in days. This variable was designated as DELAY. Each variable in the expectations, relative importance, perceptions and satisfaction sections was analyzed using the simple linear regression method with DELAY as the independent variable. Only the variable PER19 delivered a significant parameter estimate for the slope of the regression equation equal to -0.02. However, the adjusted R^2 of the procedure was less than .01, an indication of an extremely poor model fit.

The results from this test converge with those obtained in the previous sub-section that the promptness or the tardiness to reply did not influence the replies, except for reduced perception towards operating hours. However, because the model fit assessment was very poor, this result may bear little consequence.

Representativeness of the Sample

Using several methods, an examination for non-response bias revealed that all but one item were free of observable symptoms. Only variable PER19 was uncovered by both procedures. Based on the assumption that tardy respondents are more representative of the non-respondents, this raises the possibility that the sample statistics for this particular item may be biased upwards, and that the population measure would be somewhat lower than the computed estimate. However, the evidence is very weak. The regression statistics indicate a poor model fit, which make the results dubious. So, shaky evidence of biasedness on a single item is insufficient to challenge the whole sample's representativeness. Thus, there is a lack of overall evidence that early and tardy replies to survey items differ in a consistent way. In general, both early and late sub-samples can be viewed as equivalent, and the sample free from non-response bias on that basis.

Another conclusion that can be drawn from the early/tardy replies analysis is that different mixes of respondents according to some demographic attribute entailed no difference in the reply pattern. Hence, although the parameter estimates indicate that the respondents' attributes may not be represented in the sample in the same proportion as in the population, the above conclusion that there is no significant difference between the respondents and the non-respondents takes precedence. If more responses had been collected, the sample's socio-demographic profile would probably shift towards resembling more that of the population, but the responses would still be similar to those observed. So, the sample can be considered as representative.

Descriptive Statistics

A limited analysis of the descriptive statistics of the indicators is provided below. The scores and statistics of the survey items, other than those of the socio-demographic attributes, are found in Appendix D. Skewness and kurtosis are assessed at a conservative significance level ($\alpha = .001$). Computing these values as recommended by Tabachnick and Fidell (1989), the standard error for skewness for the indicators in this study is 0.124 and the standard error for kurtosis is 0.248.

Expectations, Perceptions, Differences Scores

The scores and statistics for the expectations battery and the perceptions battery are in Appendix D (Tables D1 and D2 respectively). The dimension scores were computed from the individual items and are reported in Table 4.2.

Expectations dimension scores range from 4.99 to 6.29 on the seven-point Likert scale and are positioned at values well below the maximum value of the scale. This indicates that citizens have realistic expectations towards the municipal service providers. It also supports Carlson and Schwartz's (1995) position that residents express "reasonable" requirements and do not require perfect service "in every regard" (p. 28).

In addition, this is consistent with private-sector findings on the same issue (Berry, Parasuraman, & Zeithaml, 1994). In fact, PZB's second study yielded expectations dimension scores ranging from 5.29 to 6.47 (Zeithaml et al., 1990), the latter being closer to the *ceiling* than what was observed in the current study. On one hand, this can be interpreted to mean that the realism that constrains expectations with regards to commercial services, through the market mechanism, is also prevalent for local services, although the market's influence is not so conspicuous. It would also reinforce Vehorn's (1980) proposition that a market for municipal services exists. On the other hand, perhaps customers have higher expectations towards commercial services than residents towards municipal services, exactly because consumers would be more sensitive to the resources implication of each commercial transactions than residents who may globally favour to cap tax levies by limiting services at their current level. Perhaps also commercial services are more salient to consumers, who have higher expectations because they know what they want more precisely, as compared to the sometimes-imprecise nature of some municipal services.

Perceptions dimensions scores range from 4.50 to 5.03 on the same scale. When expectations and perceptions are examined in parallel, they are almost equal for the Tangibles dimension, but the difference scores are larger for the other four dimensions. These results also coalesce with the findings of Carlson and Schwartz's (1995) that expectations scores are larger than perceptions scores and result in negative gap scores.

Gap scores were computed and are reported in Appendix D (Table D3). Table 4.2 displays the gap dimension scores. PZB's second study can be used to form an opinion on the relative ranking of the magnitude of the dimensions' gap scores. Item scores for that study are undisclosed, but a relative ranking of the dimensions gap scores results can be derived from the graphic information available in Zeithaml et al. (1990). The dimensions are ranked from one to five, with a score of one indicating the dimension that had the largest algebraic value and a score of five showing the dimension with the smallest score.

Summative gap scores for the dimensions are ranked in the same order in the current study as in PZB's. In addition, as best as can be visually determined, a similarly large discrepancy exists in the dimension scores between Tangibles and the other four dimensions, which are also dispersed within a narrower range, in PZB's study. This would first indicate that the profiles of unmet customer expectations are identical, ranking wise, in the five companies and in this municipal concern. The second similitude is that the scores for the Tangibles dimension are consistently better than those of the other dimensions. Finally, the scores for the dimensions other than Tangibles display a certain consistency among themselves, both in the commercial and the municipal concerns.

Table 4.2.

Batteries and Gap Dimension Scores, Ranking, Comparison with Another Study

Dimension	Expectations	Perceptions	Gap scores ^a	Ranking of gap-score dimension: current study	Ranking of gap-score dimension: Zeithaml et al. (1990)
Tangibles	4.99	4.88	-0.12	1	1
Reliability	6.25	4.50	-1.75	5	5
Responsiveness	6.22	4.70	-1.53	4	4
Assurance	6.29	5.03	-1.26	2	2
Empathy	6.03	4.57	-1.47	3	3

^aArithmetical inconsistencies are due to the effects of missingness and rounding.

Individuals not familiar with seven-point scaling and gap scores may experience difficulties in interpreting such results. Quiram (1995) also inferred that negative scores have an undesirable connotation. She suggested converting the metrics of the gap scores to the more familiar percentage format. In keeping with the theory, a zero value for the difference score would represent a 100% score on the transformed scale; this score is achieved when the expectations score equals the perceptions score. The maximum gap

score, -6, would translate into a 0% score. Gap scale units thus represent 16.6% increments on the transformed scale. The scores on the transformed scale would thus be computed as 100%, plus the gap score times 16.6%. For example, the score for Tangibles would be $(100 + (-.12 * 16.6))\% = 98\%$. Of course, ratings in excess of 100% are possible, occur when excessive quality is achieved, and are reflective of positive gap scores. The transformed quality scale is illustrated in Figure 4.1. This presentation is more likely to convey clearly that residents perceive that quality is high for the Tangibles dimension, while their assessment is lower for the other dimensions.

In summary, findings about the quality scale supported several antecedent findings. As compared to other studies of municipal services, the current study revealed that residents require less than optimal service and that expectations exceed perceptions on every dimension. With respect to PZB's second study, the current study revealed that Tangibles receive by far the best scores.

Importance Scores

The overall importance scale results are displayed in Appendix D (Table D4). In the aggregate, the results of the overall importance scale were compatible to those of the relative importance scale. In keeping, the Reliability and Responsiveness dimensions were respectively rated as *most important* by 40% of the respondents, and *second most important* by 37% of the respondents, while the Tangibles dimension was most frequently rated *least important* by 69% of the respondents.

Summary statistics for the relative importance of the dimensions are displayed in Appendix D (Table D5) and illustrated in Figure 4.2. The dimensions obtained unequal scores, an indication that respondents see some dimensions as more consequential than other dimensions. The Reliability dimension is awarded the highest score, Responsiveness comes second, followed by Assurance and Empathy, and last by Tangibles. The top position for Reliability supports Crompton and Mackay's (1989) finding that this dimension was the most important dimension for municipal recreation programs.

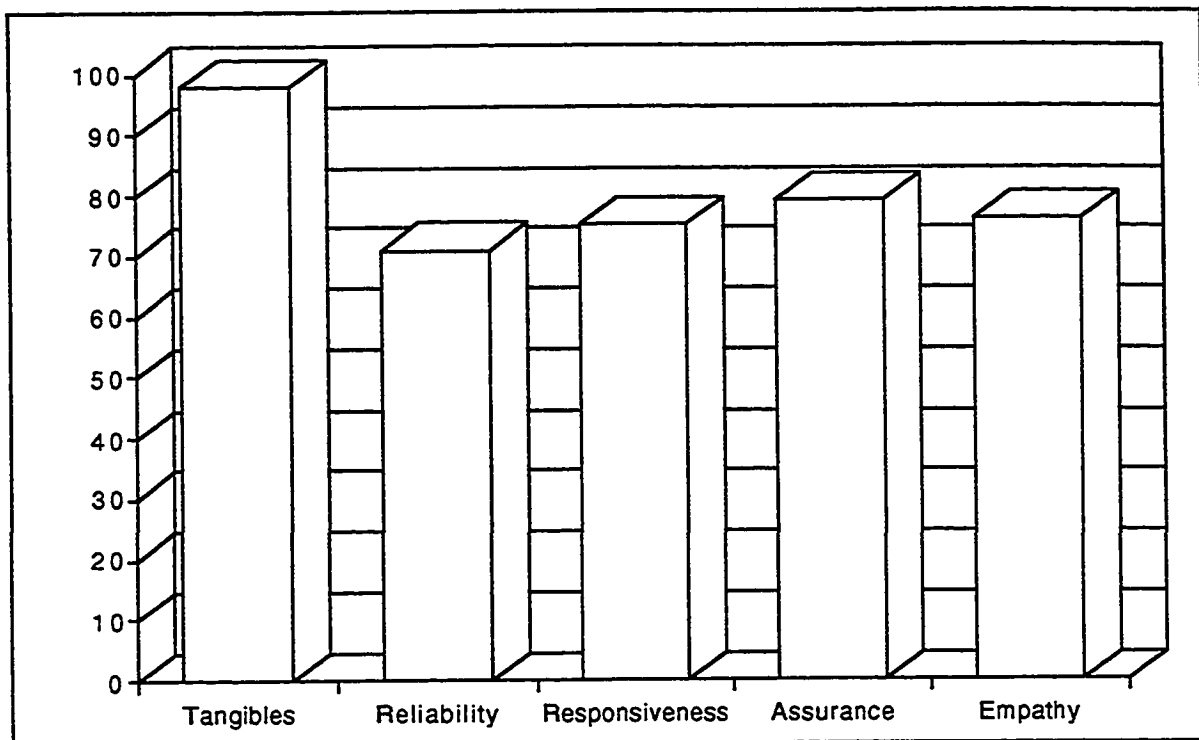


Figure 4.1. The transformed quality scale.

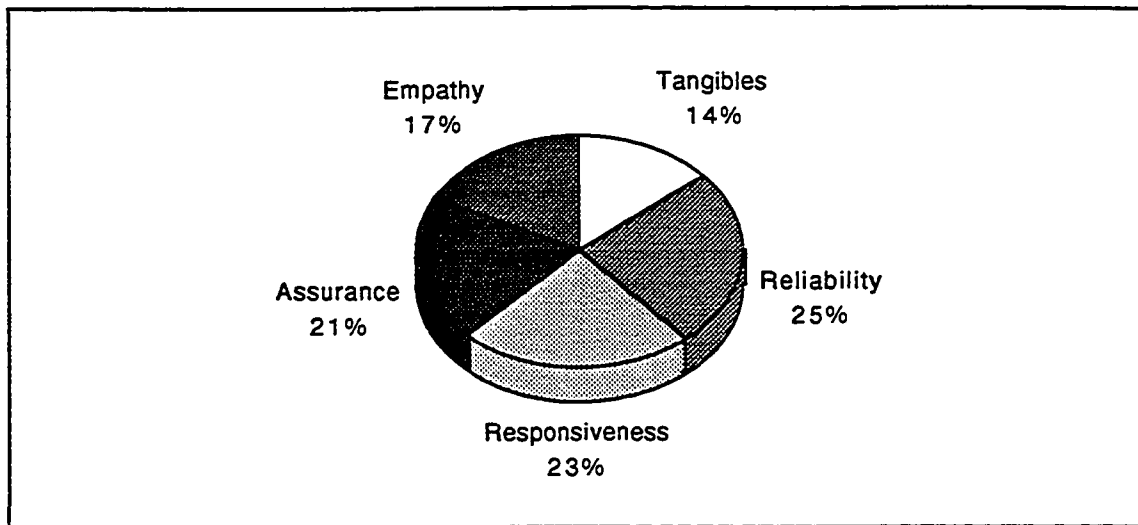


Figure 4.2. Relative importance scores.

When these scores are compared to PZB's results for the second study, more similitude between commercial and municipal services become apparent. As illustrated in Figure 4.3., the dimensions' ranking for relative importance in this study is identical to their findings. In addition, the shapes of the distributions of these scores are close. However, the Reliability dimension was graded somewhat more important for commercial than for municipal services. This is interpreted to mean that there is an added emphasis on the accuracy and dependability in the delivery of commercial services because of their salience. Concluding on this issue, PZB advocated strongly that Reliability is "the number one concern of customers today, regardless of type of service" (Zeithaml et al., 1990, p. 28).

When examined in conjunction with the difference scores, these results reveal that the dimensions that are most important to residents, Reliability and Responsiveness, are those on which the gap score is the largest in absolute value. In contrast, Tangibles, the least important dimension, enjoys the best raw quality score. Thus, it appears that there is an inverse relationship, although imperfect, between the relative importance of a dimension and the raw quality score it achieves. A similar pattern was observed by Carlson and Schwartz's (1995).

In summary, Reliability is the most important dimension and there is a somewhat inverse relationship between the dimension's perceived quality score and their relative importance, here as well as in many previous studies.

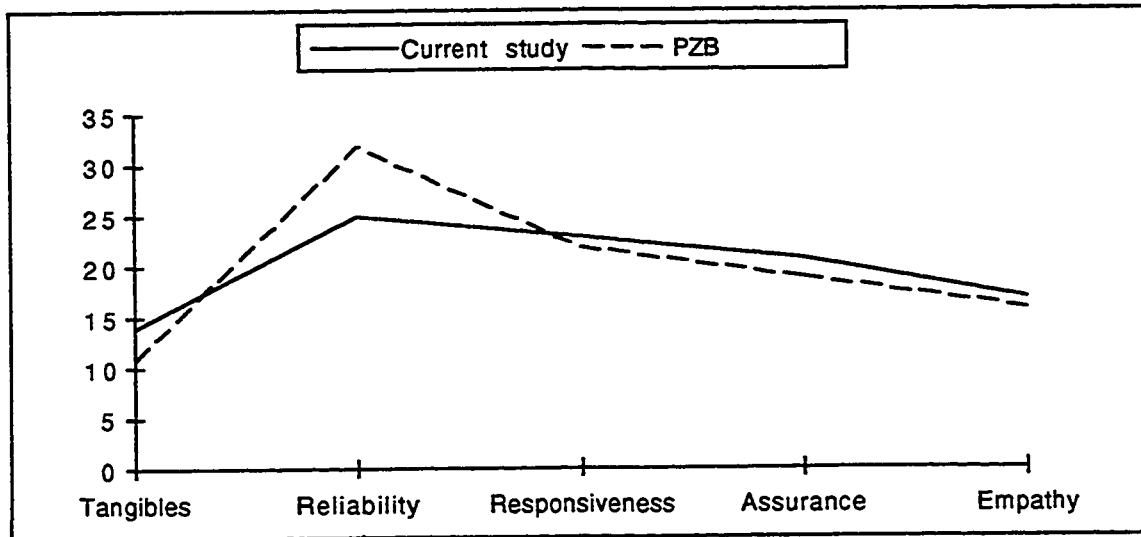


Figure 4.3. Comparison of the relative importance scores for the current study and PZB's study⁸.

⁸ Zeithaml et al., 1990, p. 28.

Satisfaction Scores

Satisfaction scores are found in Appendix D (Table D6). The scores for the two negatively worded items were reflected to ensure that the measures were unidirectional. The results obtained coalesce with those of Peterson and Wilson (1992) who found that self-reported satisfaction surveys generally show moderate to high positive levels of satisfaction and negative skewness. However, Peterson and Wilson's (1992) findings that positive items have higher scores than negative items was not supported in this study, with the negative items exhibiting an average score of 5.37, while the positively worded items displayed an average score of 4.65. Because no previous study on municipal services also using SERVQUAL had reported independently measured satisfaction scale, no parallel can be drawn on this issue.

Preparation of the Data for Multivariate Analysis

In this sub-section, the issue of missing data and its resolution will first be examined. Next, the issue of the suitability of the data for multivariate analysis will be addressed. Finally, the data will be screened for outliers.

Missing Data

Missing data occurred in 100 of the 391 cases. Deleting seven cases that displayed the worst missingness resolved about half of the missing data points, reducing the number of missing occurrences to 177. The remaining 384 cases were deemed usable for analysis.

Tabachnick and Fidell (1989) asserted that small amounts of missingness carry little consequences. Furthermore, these authors added that a pattern of occurrence of the missing data is consequential. Forty-three variables displayed five or fewer missing data points, too few points to identify a missingness pattern. Consequently, the missing data in these variables is assumed to be missing completely at random (MCAR). This form of missingness is ignorable (Little & Rubin, 1987).

Seven variables exhibited moderate levels of missingness. These variables are (a) IMPMOS, (b) IMPSEC, (c) IMPLEA, (d) PER9, (e) SAT1, (f) SAT4, and (g) INCO. The three overall importance variables showed a similar pattern of missingness because the items may have been more difficult to complete. This intrinsic difficulty level could cause the absent replies not to be missing at random (NMR). However, because these variables have exhausted their usefulness for statistical analysis, they were disregarded.

The scores for survey items considered as interval were examined for the effect of missingness in variables PER9, SAT1, and SAT4 through *t-tests*. When verifying the scores of ordinal variables, the Mann-Whitney-Wilcoxon rank sum nonparametric test was used. For all tests, significance was tested at the threshold of $p = .05$ level. No important pattern of missingness could be identified for these three variables.

Variable INCO was examined in a similar manner and showed a pattern for six items. Items on questionnaires eliciting the income level of respondents have been documented to show high missingness levels, with the missingness attributed to resistance to provide the information, sometimes in association with old age. This would suggest that the missingness in the INCO variable could be assumed to follow this historical pattern and would not be missing at random (NMR). However, because it represents only 13 occurrences, it appears reasonable and efficient to overlook it. Consequently, all missingness was considered as ignorable.

Values were imputed for missing data points to maximize the number of complete cases available for analysis. A composite procedure, based on the regression and the hot-deck imputation techniques and mentioned by Little and Rubin (1987), was used for imputation. According to these authors, the conditional mean imputation—regression imputation using a single regressor—provided “the best point estimates of the missing values” (Little & Rubin, 1987, p. 61), but introduced bias due to variability restrictions. Instead, they suggested that “imputations are selected randomly from a distribution of plausible values, rather than from the centre of the distribution. One way of achieving this

is to add a suitable perturbation to the conditional mean" (Little & Rubin, 1987, p. 47). Rounding was thus introduced in the filling-in process to provide a residual. Because all rounded values are within the range of observed values, the second step falls within the hot-deck family of procedures (Little & Rubin, 1987, p. 62). In addition to its methodological advantages, this method was also retained because it preserves the properties of the distribution of the variables.

Consequently, the imputed value was the rounded conditional mean value obtained from the variable with the highest correlation to the variable displaying missingness, within the dimension. For example, variable EXP1 was missing for case #260. The variable within the Tangibles dimension that displayed the highest correlation to EXP1 was EXP2. The value of EXP2 for the case was seven. The mean value of EXP1 for complete cases displaying a score of seven for EXP2 was computed to be 5.27, and was rounded to five, a value also observed for 108 other cases. Thus, a five was filled-in as the value for EXP1 in case #260. Other cases and variables were treated alike. Means, standard deviations, and correlations for the data set after imputation for missingness are provided in Appendix E.

An early estimation of the simultaneous model by *CFA*, using the listwise deletion data set ($N = 324$) and the replaced data set ($N = 384$), revealed that there was no important differences in parameter and fit estimates between the data sets. This constituted a clue that the replacement mechanism had not introduced bias into the data. The replaced data set was used thereafter in view of its size.

Appropriateness of the Data Set for Factor Analysis

Suitability of the data for factor analysis can be assessed by jointly examining several data attributes. These attributes are (a) sample size, (b) absence of multicollinearity, (c) the covariance matrix is not an identity matrix, and (d) variables *fit* together.

The likelihood of obtaining convergence and a proper solution during optimization is directly related to the size of the sample; a problem is not likely to occur when there are 300 cases or more (Anderson & Gerbing, 1984). In addition, under Comrey's (1973)

classification scheme for sample sizes, this sample can be assessed from *good* to *very good*.

The size of the sample is also relevant when examined in perspective with the number of free parameters being estimated. A ratio of cases to free parameters (C/P) of five to one is adequate when using a normal theory based estimator such as ML (Bentler & Chou, 1988, p 173). At most, models under scrutiny will include 188 free parameters, for a C/P ratio of 2:1, which is short of the comfort zone. However, such a large model is an extreme situation. The remainder of the models being estimated enjoyed a C/P ratio that is larger, closer to the suggested threshold, and more likely to yield proper solutions. When the model is too large, a solution may sometimes be obtained but not robust statistics, due to additional computational requirements unfulfilled in view of the limited sample size.

Multicollinearity exists when at least one variable is linearly dependent on another variable. Should this condition prevail, the matrix becomes singular and cannot be inverted (Morrison, 1976, p. 47). The ML estimation method requires that “this matrix be positive definite and invertible” (Bentler & Chou, 1988, p. 184); hence, a singular matrix is ill conditioned for factor analysis with the ML estimator used for the current study. The determinant ($|A|$) of a non-singular matrix is significantly greater than zero (Morrison, 1976). Computing the determinant of the correlation matrix of the variables, a procedure that is automated in EQS, can be used to assess this. The value for $|A|$ is 0.0037642. Although the value is small, it fulfils the necessary condition for admissibility of the input matrix.

The next test of suitability for factor analysis is Bartlett’s (1954) test of sphericity, which assesses if the off-diagonal elements of the correlation matrix are zeros. The test result is 13012.94 (p -value < .001). A small significance level implies the rejection of the null hypothesis and suitability of the data set. The hypothesis that the covariance matrix is an identity matrix is thus rejected.

The last test of suitability is the Kaiser-Meyer-Olkin measure of sampling adequacy (KMO), which tests the hypothesis that partial correlation values in the sample are small (Tabachnick & Fidell, 1989, p. 604). Larger values for the index indicate that the variables belong together and are fit for factor analysis. The value for this statistic is .92, qualifying the sample as “marvelous” (Norusis, 1988, p. 129).

In summary, the sample size can be qualified as good or very good although the C/P ratio is low. The input matrix is positive-definite, amiable to the ML estimator. The hypothesis that the input matrix is a null matrix was rejected. Finally, the data set has a marvellous fit. Hence, the data available for analysis in this study is most appropriate for factor analysis.

Outliers

The presence of outliers in the data set may have a detrimental effect on the analyses. Bentler and Chou (1988) asserted that “a few extreme outliers are unlikely to be described by a structural model that describes all the remaining observations” (p. 166).

However, inherent variability reflects the genuine “amount of variability indigenous to the population” (Barnett & Lewis, 1994, p. 33) and although discordant, delinquent values belong to the population. Reducing the variability of the sample by removing them conflicts with the randomness of the sample, which becomes censored (Barnett & Lewis, 1994, p. 40). Thus, extreme outliers should be considered as nuisances and “robust methods of analysis to minimize their impact” should be employed (Barnett & Lewis, 1994, p. 39), which is the case in the current study. Consequently, all observations have been retained for analysis.

Model Estimation

The steps that were followed for model estimation are (a) evaluation of model identification, (b) model assessment, and (c) possible model re-specifications.

Evaluation of Model Identification

The model of interest has a parameter fixed to assign a scale to each factor, it is recursive, and is parsimonious as will be evidenced by the good scores obtained on parsimony fit indices. Thus, it fulfils the prevention conditions suggested by Schumacker and Lomax (1996). The availability of degrees of freedom for estimation demonstrates that it also meets the necessary order condition. In addition, both empirical identification tests performed by the EQS software indicate that the sufficient condition for identification is also fulfilled. These tests converge to demonstrate that the model is identified.

Model Assessment

A baseline version, two measurement versions, as well as a structural version of the simultaneous model are estimated. The EQS software syntax for the estimation of the structural model is attached in Appendix F. A selection of computed fit indices related to these models is displayed in Table 4.3. Estimated parameters are discussed for each model, but only the standardized estimates for the structural model are reported. The baseline model is used only as a basis to compute some fit indices; hence, it is not discussed further.

Eleven-Factor Measurement Model

The 11-factor measurement model was estimated successfully, but robust statistics are not available. The acquisition of robust statistics are “computationally very demanding” (Bentler, 1995, p. 47). The concurrence of a very large model and a not-that-large data set has precluded their availability. Thus, this model must be examined strictly through other tools, such as residuals and component measures.

Table 4.3.
Fit Indices for the Alternate Models

Model	\bar{T}	df	GFI	RMSEA*	CFI*	NC*	PCFI*	$\Delta \bar{T}$	Δ df
1. Eleven-factor correlated measurement model	--	1050	.76	--	--	--	--		
2. Second-order measurement model	1856.78	1092	.74	.04	.90	1.70	.84		
3. Simultaneous model	1856.50	1092	.74	.04	.90	1.70	.84		
Difference between Model 3 and Model 2								-0.28 ^{††}	0
4. Baseline model	8789.15	1176	--	--	--	--	--		

Note. \bar{T} = Satorra-Bentler robust T statistic; GFI = goodness-of-fit statistic; RMSEA* = robust root mean square error of approximation; CFI* = robust comparative fit index; NC* = robust normed chi-square index; PCFI* = parsimonious comparative fit index; $\Delta \bar{T}$ = likelihood ratio test. Dashes (--) indicate that no value was obtained for this model or test. [†] $p < .05$, one-tailed; ^{††} $p < .01$, one tailed.

The average absolute standardized residual of the model is .04. Bollen (1989) mentioned that standardized residuals can be interpreted more easily than their raw counterparts because they "provide an approximate correction for . . . sample size effects and for scaling differences" (p. 259). The displayed values are small and manifest a generally close fit. Variable PER3, however, produces about half of the top 20 largest paired-variable residuals (maximum value $\leq .29$) and suggests a possible problem.

The factor loading estimates are all significant ($p < .01$). The standardized factor loading estimates range from .33 to .90. The strength of the relationship between the factors and their respective indicators is supportive of the theory. Inter-factor correlation values within each battery are all significant ($p < .01$). This high rapport between the dimensions suggests that the theory-driven underlying second-order constructs of expected quality and perceived quality might be present.

Second-Order Factors Measurement Model

The two second-order factors for quality and the factor for satisfaction linked obliquely form the actual measurement model, which was estimated next. The average

absolute standardized residual of this model is .05, larger than that of the 11-factor model, with the PER3 variable still prone to large residuals. The \bar{T} statistic is significant, which implies a misfit. However, the T index “has a tendency to [unduly] indicate a significant probability level” for sample sizes over 200 (Schumacker & Lomax, 1996, p. 125). The GFI index also yields a low value. Both these indices would suggest rejecting the model. Other indices, however, support the hypothesis that the measurement model is adequate. The CFI* index meets the .90 threshold generally recommended for normed indices (Mueller, 1995). The RMSEA* index value is .04, indicating a *close* fit, as suggested by Browne and Cudeck (1993). The NC* index yields the good score of 1.70, because values as large as 5.0 have been seen as indicator of acceptable fit (Mueller, 1995). The PCFI* score of .84 is satisfactory because Rigdon (1996) suggested a score of .75 or above as indicative of model parsimony. In summary, some of the indices of model fit suggest a good model fit, while others do not, but the bulk of the evidence is that the measurement model is adequate.

Also of interest, the inter-factor correlation values, displayed in Figure 4.4, indicate that only the perceptions and satisfaction factors correlate in a large manner, hinting that perhaps the expectations may not constitute an important part of the model.

Structural Model

The structural version of the simultaneous model was estimated successfully. The average absolute standardized residual of the model is .05, similar to the measurement model. The standardized measurement and structural parameters estimates for the structural model are attached in Appendix G. All the factor loading estimates are positive and significant ($p\text{-value} < .05$). The standardized factor loading parameter estimates are generally large. The standardized structural parameter estimates are also large and positive. The coefficients of determination values for the first to second-order factors confirm that most first-order constructs make an important contribution to their respective higher level factors.

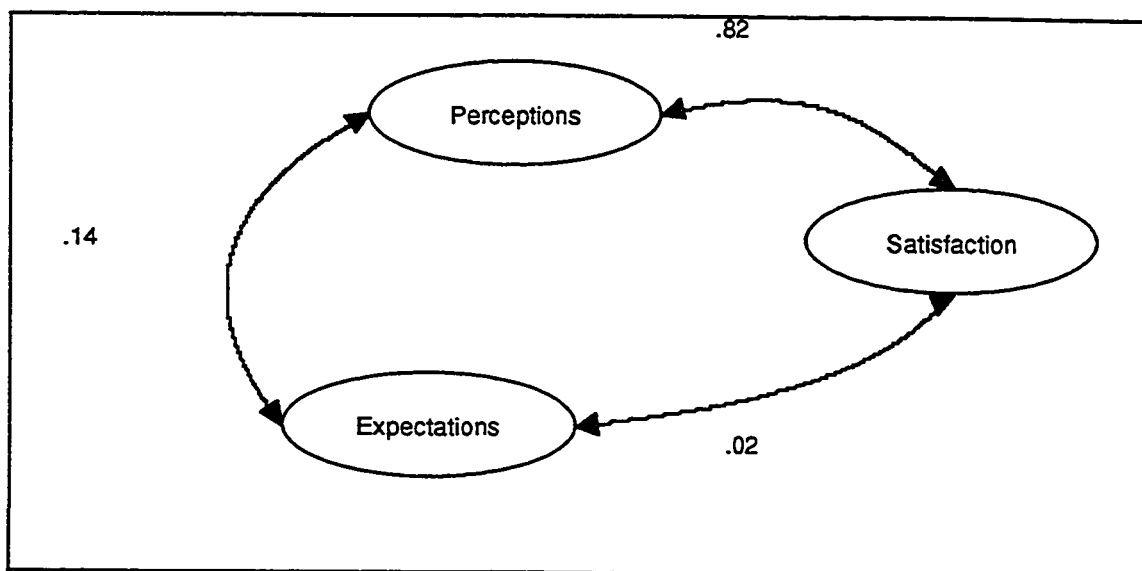


Figure 4.4. Partial standardized estimates of the measurement parameters for the simultaneous model.

The specification of the structural relationships has not resulted in any degradation of fit. Because the second-order factor measurement model and the structural model have the same number of degrees of freedom, the likelihood ratio test cannot be performed. However, the identical results on fit indices manifest that the replacement of free covariance links by directionality specified relationships has not hampered the fit. So, the bulk of the evidence supports the hypothesis that the structural model is a good representation of the underlying concepts, given the current data set.

As can be observed in Figure 4.5, expectations contribute only mildly to the shaping of perceptions; moreover the R^2 of this particular equation is .02, indicating that this regression equation is almost trivial in explaining the variance that occurs between these factors. Nonetheless, expectations seem to exercise a modest but significant negative influence on the shaping of satisfaction. Perceptions, however, have a more determinant influence on satisfaction. The regression equation linking the two determinant factors to satisfaction enjoys an R^2 of .68, which indicates that the bulk of the variance between the structural variables is explained successfully by the model.

Re-Specification

While the model of interest has demonstrated a close fit, possible re-specifications have been examined. These *post hoc* model modifications were suggested by (a) the behaviour of a case, (b) the residuals analysis, (c) the Wald test on free parameters, (d) the LM test on fixed parameters, and (e) the theory-driven perception-only paradigm. These alternatives are examined below and results of their estimation are reported in Table 4.4.

Case Removed

The first alternative model tests the simultaneous structural model, but removes case number 216 that showed the largest kurtosis deviation in the sample. This alternate model was estimated and resulted in degraded indices, as well as an increase in the \bar{T} value with the same number of degrees of freedom, all of which are indicative that this avenue is not fruitful.

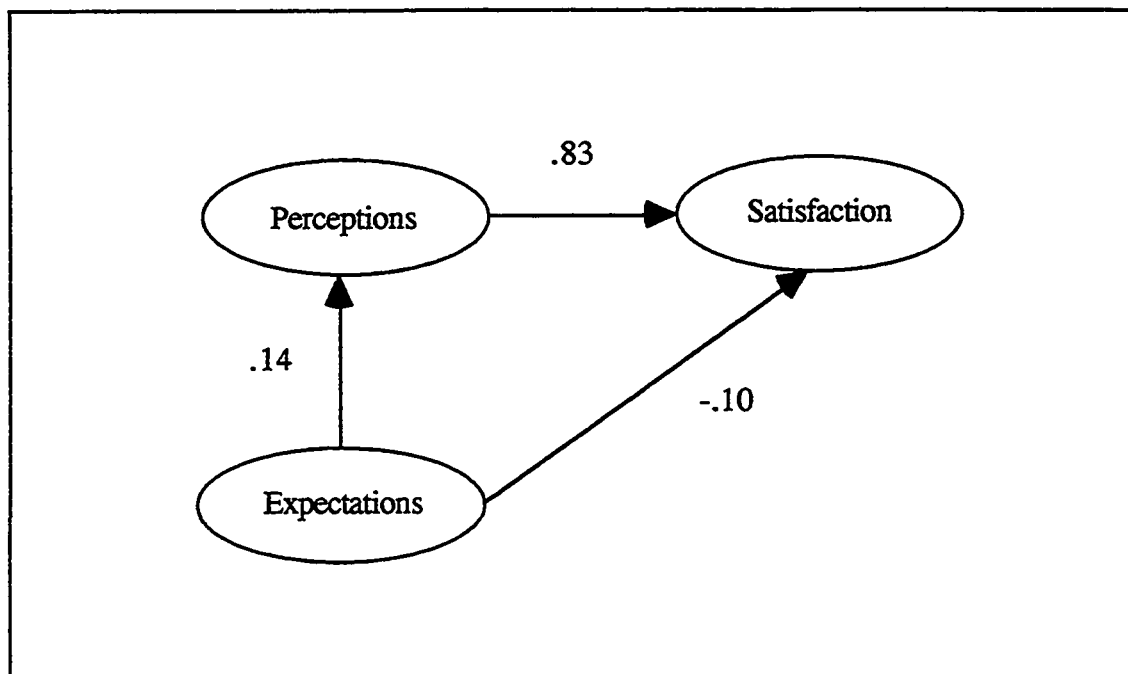


Figure 4.5. Partial standardized estimates of the structural parameters for the simultaneous model.

Table 4.4.
Fit Indices for the Re-Specified Models

Model	\bar{T}	df	GFI	RMSEA*	CFI*	NC*	PCFI*	$\Delta \bar{T}$	Δ df
5. Remove case=216 model	1927.31	1092	.74	.05	.89	1.77	.83		
Difference between Model 5 and Model 3								70.81 ^{††}	0
6. Remove EXP3&PER3 model	1694.14	1000	.75	.04	.91	1.69	.77		
Difference between Model 6 and Model 3								-162.36 ^{††}	-92
7. Remove correlated errors model	1925.61	1114	.73	.04	.89	1.73	.85		
Difference between Model 7 and Model 3								69.10 ^{††}	22
8. Cross-loadings for variables EXP21&PER21 model	1824.15	1090	.75	.04	.90	1.67	.84		
Difference between Model 8 and Model 3								-32.35 ^{††}	-2
9. Perception-only model	727.12	318	.78	.06	.93	2.29	.84 ^a		
Difference between Model 9 and Model 3								-1129.38 ^{††}	-774

Note. \bar{T} = Satorra-Bentler robust T statistic; GFI = goodness-of-fit statistic; RMSEA* = robust root mean square error of approximation; CFI* = robust comparative fit index; NC* = robust normed chi-square index; PCFI* = parsimonious comparative fit index; $\Delta \bar{T}$ = likelihood ratio test.

^aBaseline with 351 degrees of freedom.

[†] $p < .05$, one-tailed; ^{††} $p < .01$, one tailed.

The questionnaire and the socio-demographic profile of the respondent were examined to uncover clues that this is an outlying observation. The respondents' expectations were generally high, but not always extreme. This person is neither old nor uneducated, which makes misunderstanding of the items wording unlikely. Thus, although there is weak evidence that this case could be an outlier, data contamination was not established. There is no substantive reason to remove this case and no practical advantage in doing so. Thus, this first alternate model is rejected.

Variables Removed

The second alternate model deals with the issue of variable PER3. Because this variable demonstrated misfit with several other indicators, it was deleted, along with the parallel EXP3 variable, and a model testing a reduced set of indicators was estimated. This reduced model resulted in a generally improved set of fit indices, with mixed evidence concerning parsimony. The likelihood ratio test yielded a significant result, an indication that this alternate model may fit better than the model of interest.

Variable PER3 concerned the appearance of the employees. It may have contained some ambiguity, because municipal governments retain the services of both uniformed employees such as firemen for whom the appearance may be an observable phenomenon, as well as non-uniformed employees with respect to whom the issue may be trivial. Carman (1990) found that the dress item was not critical in all settings, but kept the item nonetheless. However, this indicator is part of the Tangibles dimension, which provides the least contribution to the higher order concepts. Hence, the distinction between the variable's presence and absence may be less critical than the likelihood ratio test seems to suggest. Consequently, the conservative approach motivates keeping the variable in view of its theoretical founding, although the likelihood ratio test results suggest differently. Hence, it appears appropriate to reject this alternate model on a theoretical basis.

Correlated Errors Removed

The results of the Wald test indicated that it might be beneficial to remove a few of the correlated errors. Because the correlated errors were hypothesized to be part of the model for a substantive reason, it was not logical to remove only some of them. Consequently, the third alternative was a more restricted model with all correlated errors removed. Its estimation resulted in a somewhat degraded fit except for parsimony indices, which improved slightly. The significant likelihood ratio test suggests rejecting the hypothesis that errors are not correlated because the gain in degrees of freedom leads to a significant increase in the magnitude of the \bar{T} index.

Cross Loading Links Added

The LM test suggested freeing the path between the dimensions and several indicators theorized to load on other factors. Principally, the double loading estimates involved the Reliability dimension. So, a fourth alternate model allowed for the cross loading of variables EXP21 and PER21 to both the Empathy and the Reliability dimensions. This resulted in a general improvement of fit indices. Further, the likelihood ratio test yielded a significant result ($p\text{-value} < .01$), again an indication that the alternate model may be better than the model of interest. However, from a theoretical stance, moving away from a simple structure leads to ambiguous interpretation. Baumgartner and Homburg (1996) cautioned against the introduction of parameters “simply to boost the fit of the model” (p. 146). So, although this alternate is empirically supported, it is believed that this may be only a sample effect that might not replicate in another data set. Thus, this alternate model is rejected on substantial grounds.

Perception-Only Model

The additional matter to investigate stems from the literature. It verifies the hypothesis that the perceived quality measures alone define a satisfactory predictive instrument. To assess this hypothesis, a reduced model based on the simultaneous model but leaving out the expectations battery and its structural links was estimated. This model shows a considerable improvement on fit indices, except on parsimony instruments where the evidence is mixed. The likelihood ratio test shows a significant improvement over the model of interest. The remaining standardized first/second-order factor loading parameters estimates, displayed in Appendix H, are generally quite similar to those of the model of interest. Most differences are only a few marks at the second decimal level. Basically, the factor pattern as well as the magnitude of its links remains unchanged.

When examined from a measurement perspective, the perception-only model is formed of a second-order factor for perceptions linked obliquely with the satisfaction factor. When this measurement model is estimated, it yields a \bar{T} statistic value of 727.17

with 318 degrees of freedom. The likelihood ratio test value between the scores of these measurement and the structural model is negligible with no degrees of freedom. This evidences the appropriateness of this reduced structural model.

The causal link and its estimate are illustrated in Figure 4.6. There is a negligible difference between the models with regards to the estimated strength of the causal link between perceived quality and satisfaction. Consequently, the estimated outcome should be invariant when computed with either method and the loss of information resulting from the absence of the expectations battery lacks computational consequences. We may then conclude that the perception-only model is a better fitting and more parsimonious alternative to the simultaneous model.

In conclusion, the model of interest, comprising two second-order factors causally linked between them and to the satisfaction dimension, was supported by the data through a good fit. However, worthwhile improvements through *post hoc* model modifications suggested that the perception-only model would be a preferred alternative and is retained over the model of interest. The next step is to assess the psychometric properties of the scales comprised in the reduced model.

Psychometric Properties of the Scales

To assess the psychometric properties of the scales of the perception-only model, the data was analyzed through a six single-dimension factor pattern—one for each of the five dimensions for the perceived quality battery and one for the satisfaction scale. Then, these factors were subjected to validity and reliability tests.

Factor Pattern

All factor loading estimates are significant ($p < .01$). The standardized loading estimates of the indicators on the unidimensional factors are displayed in Appendix I; most of these estimates exceed the .70 value recommended by Segars (1997).

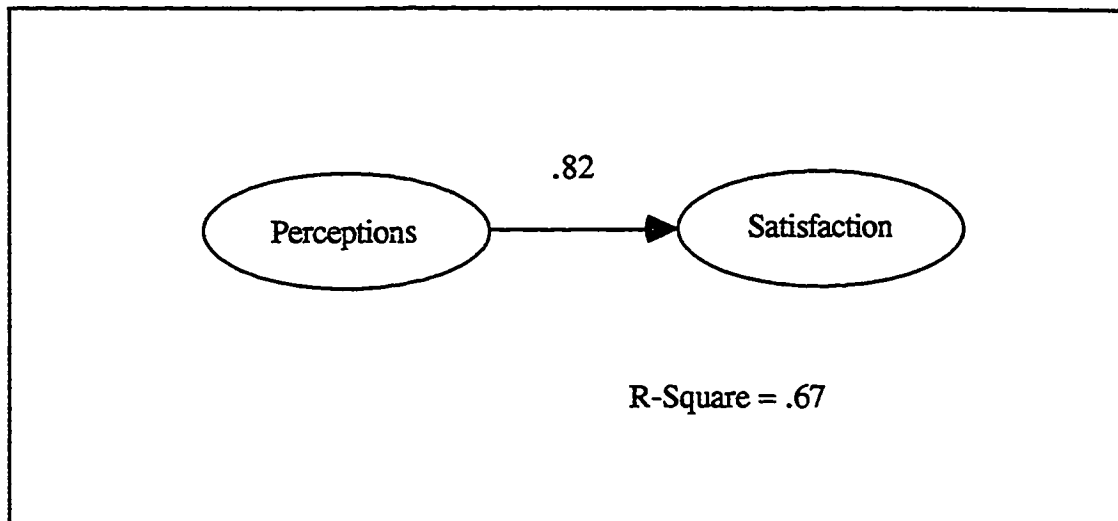


Figure 4.6. Partial standardized estimates of the structural parameters for the perception-only model.

Fit indices are provided in Table 4.5. The usual parsimony fit indices estimates are not reported because the small size of these models makes these estimates not warranted. The RMSEA* statistic is not provided either because most estimates yielded abnormally high results due to the very small number of degrees of freedom for each model. For the perceptions battery, all the dimensions achieved the usual fit threshold for GFI and CFI*, with mixed results for the Responsiveness dimension. The fit of the satisfaction scale is adequate.

Table 4.5.

Fit Indices for the Unifactorial Dimensions

	\bar{T}	df	GFI	CFI*	α
Tangibles	5.51	2	.99	.98	.70
Reliability	37.84 ^{††}	5	.92	.96	.89
Responsiveness	26.04 ^{††}	2	.87	.97	.90
Assurance	15.51 ^{††}	2	.95	.98	.89
Empathy	30.69 ^{††}	5	.93	.97	.89
Satisfaction	21.37 ^{††}	5	.97	.97	.79

Note. \bar{T} = Satorra-Bentler robust T statistic; GFI = goodness-of-fit index; CFI* = robust comparative fit index.

^{††} $p < .01$, one-tailed.

Content Validity

The content validity of the SERVQUAL scale has been examined in the previously reported studies. There is a prevalence in the literature that the 22-item instrument possesses this attribute (Cronin & Taylor, 1992; Parasuraman et al., 1991a; Parasuraman et al., 1988), although some authors do not agree (Johnston et al., 1990). Consequently, the content validity of the SERVQUAL items need not be demonstrated again because this form of validity, including the trait validity component, was manifest to the authors of the scale as well as some others who examined the matter explicitly. The content validity of the satisfaction scale arises from the use of a majority of seasoned items that proved their worth in previous applications.

Criterion-Related Validity

Criterion-related validity could be examined from a concurrent or a predictive approach. The difficulty arising from the concurrent approach is that an array of criterion measures would yield different results, casting doubts as to which is the proper measure. Furthermore, including many possible criterion items becomes a self-fulfilling prophecy. Hence, due to the lack of practical utility of this test, no such items were included in this survey for this purpose. In the predictive context, the ability of quality to predict satisfaction constitutes the prime object of one of the research questions, which will be tackled subsequently to the distinct examination of the concepts. Hence, testing for both satisfaction as a criterion for quality as well as quality as a predictor of satisfaction appears redundant, if appropriate at all. This discussion reinforces the earlier statement that this form of validity appears useless as an empirical assessment tool. Consequently, criterion-related validity of the scales will not be examined.

Construct Validity

The construct validity of a set of scales is demonstrated when each scale possesses (a) unidimensionality, (b) reliability, (c) convergent validity, and (d) discriminant validity.

Unidimensionality

The unidimensionality of the scales is assessed by examining the fit indices of the different dimensions. Possible misfits are indicated by the large values for the \bar{T} statistic, but these will be discounted for the previously mentioned reasons. The other fit indices display values well in excess of the .90 threshold, indicative of a good fit of the single factor dimension models, for all but the Responsiveness dimension, for which the evidence is mixed. Thus, the scales appear unidimensional in general.

Reliability

All scale reliabilities exceed largely the .50 level for α recommended for exploratory studies such as this one and most meet the .80 level recommended in general by Nunnally (1967).

Convergent Validity

The three criteria on which convergent validity is assessed are (a) the significance of the factor loading estimates, (b) the magnitude of these estimates, and (c) the value of the CFI* index.

All loading estimates are significant. In magnitude, they can be qualified of mid to almost large, but not impressive. All scales enjoy a respectable score on the CFI* index. Consequently, the convergent validity of the scales is demonstrated.

Discriminant Validity

Testing for discriminant validity consists of a flurry of tests comparing the dimensions as a freely correlated pair forming a two-factor model to an alternate single factor model. The assessment is made through likelihood ratio tests, with the difference (Δ) between the \bar{T} statistic for each model distributed as a χ^2 , with a single degree of freedom.

For this study, the method suggested by Anderson and Gerbing (1988) was used to select the individual significance levels. Because an overall significance level of $p = .05$ was desired, the individual test significance was set at the stringent level of $\alpha_t = .005$. The $\Delta \bar{T}$ statistic must therefore be equal to or larger than 7.88 to achieve significance. The results are displayed in Table 4.6. Despite a rather stringent method of computing the threshold, all scales demonstrated their discriminant properties.

Table 4.6.

Likelihood Ratio Paired Test Values for Unifactorial Dimensions

	Tangibles	Reliability	Responsiveness	Assurance	Empathy	Satisfaction
Tangibles						
Reliability	86.16					
Responsiveness	73.22	69.55				
Assurance	65.37	90.98	13.72			
Empathy	87.65	147.83	91.03	38.71		
Satisfaction	114.70	124.05	164.61	117.29	133.25	

Note. All values non-significant ($p > .05$, one-tailed).

Summary Concerning the Psychometric Properties of the Scales

The bulk of the evidence permits to conclude that the psychometric properties of the perceived quality and satisfaction scales hold in general. Content validity was ascertained for all scales. Criterion validity was not assessed. The data yield unidimensional scales, although a dimension fails to achieve a high enough score on the GOF statistics. Reliabilities of the scales are good. On the issue of convergent validity, the data yields significant but not that large loading estimates, while all dimensions show a sufficient achievement on the CFI* statistic. On a rash test of discriminant validity, all dimensions distinguish themselves clearly from the others. In conclusion, the scales are reliable and the bulk of the evidence is in favour of affirming the validity of the scales.

Testing of the Research Questions

In this sub-section, we will examine the results of the tests performed to investigate the research questions, along with their interpretation.

Research Question 1

The first research question is concerned with the ability of the data to support the assumptions of the perception-only model to adequately represent the quality of municipal services.

The above-mentioned assessment of the model fit and of the psychometric properties of the scales yielded satisfactory results for the five quality scales. However, the factor pattern of SERVQUAL is a debated issue. A large number of studies found SERVQUAL to be unidimensional, which is not the case in this research. This issue is worthy of further investigation. Because many studies used *EFA* as an assessment mechanism, this procedure was used next to provide comparable results. The results of the alternate models assessed using *EFA* are summarized in Table 4.7.

When the solution was constrained by the eigenvalues ≥ 1 rule, four factors were retained. The magnitude of the eigenvalues were 13.46, 1.64, 1.41, and 1.24. The cumulative variance extracted was 60%. When obliquely rotated, Tangibles loaded on one factor, but the remainder pattern matrix was not meaningfully interpretable.

Using the scree plot decision rule suggested two factors. When the factor pattern matrix was extracted and rotated, most of the quality and satisfaction items loaded on a single factor, while the second factor behaved as a contrast.

To replicate the unidimensional findings of previous studies, a single factor solution was extracted. All but one item displayed loading estimates of .30 or more and the factor accounted for about half of the variance.

Finally, a six-factor theory driven solution was also extracted and rotated. The last two eigenvalues had magnitudes of 0.98 and 0.85, respectively. The interpretation of the factors was also deficient, but closer to the anticipated results. Satisfaction loaded on one factor, Tangibles on another one, Reliability on a third with some overlap on the remaining dimensions, two factors were loaded with mixed items, and the sixth factor was a negatively loaded contrast.

Table 4.7.

Exploratory Factor Analysis Results

<i>N</i> of factors	Criteria	Cumulative variance extracted	Interpretation
1	Previous studies	48%	Single factor
2	Scree plot	52%	Single factor
4	Eigenvalues ≥ 1	60%	One dimension identified
6	<i>A priori</i>	63%	Three dimensions identified

Thus, the *EFA* did not appropriately represent the expected factor pattern. In fact, the use of this procedure does not really shed much light on the factor pattern in the data. Marsh (1987) mentioned that the advantages of *CFA* over *EFA* “are especially important for the examining of hierarchical structures”. When confronted with such complex models, perhaps *EFA* results are blurred by the overlapping single higher order factor and strongly correlated multiple first-order factors. The outcome of this line of research is indicative of the superiority of the *CFA* to ferret second-order factor models.

These results also provide a framework to interpret the prevalence of one and five factor solutions in the literature. Because the *true* relationship between the factors is a hierarchical one, the complex nature of the model and the limits of the assessment instrument provoked the erratic results. One factor was found when the top factor had more empirical prevalence, while five factors were identified when the correlated first-order factors were better defined in the data. The hierarchical structure was always present, but researchers either saw one or the other, the result of a sample effect. The mechanism of this framework is evidenced in Babakus and Mangold’s (1989) finding of a unidimensional scale when using disaggregated measures, while confirming the second-order factor model by using partially aggregated dimension scores.

To rule out the hypothesis of a unidimensional quality factor pattern, a modified perception-only model was first estimated by *CFA* with only two factors, one for quality and one for satisfaction (model 10). As can be observed in Table 4.8., this model is rejected because of an obvious degradation of the goodness-of-fit indices and poor incremental fit test results.

Because the Tangibles dimension yielded lower loading estimates compared to the other quality dimensions, another model was developed, comprising satisfaction and a second-order factor model, the latter formed of two first-order factors, Tangibles and another factor that encompassed the other four hypothesized dimensions (model 11). The Tangibles and others model is also rejected in view of the degraded fit and the poor results

on the likelihood ratio test. From this additional array of tests, we can conclude that the perception-only model is the best available representation of the theory that is supported by the data.

Table 4.8.

Fit Indices for the Basic Structure Models

Model	\bar{T}	df	GFI	RMSEA*	CFI*	NC*	PCFI*	$\Delta \bar{T}$	Δ df
10. Two factors model	1094.45	323	.71	.08	.87	3.39	.80		
Difference between Model 10 and Model 9								367.33 ^{††}	5
11. Tangibles & others model	1008.55	321	.72	.07	.88	3.14	.80		
Difference between Model 11 and Model 9								281.43 ^{††}	3

Note. \bar{T} = Satorra-Bentler robust T statistic; GFI = goodness-of-fit statistic; RMSEA* = robust root mean square error of approximation; CFI* = robust comparative fit index; NC* = robust normed chi-square index; PCFI* = parsimonious comparative fit index; $\Delta \bar{T}$ = likelihood ratio test.

[†] $p < .05$, one-tailed; ^{††} $p < .01$, one tailed.

Consequently, the answer to the first research question is that the data supports the hypothesis that the second-order factor model linking the perceived quality dimensions is a valid and reliable model to represent the quality of municipal services. This finding has several practical and methodological implications.

First, from a managerial perspective, this instrument constitutes a valuable tool for city administrators and elected officials with which to gauge the quality of their services from the residents' perspective. This fills a gap in the array of available managerial tools and enables local public deciders to dedicate more attention to the perceived quality of their services. Data collected through this instrument can now be examined in perspective with objective measures of service quality, such as response time, attendance, and activity level, to isolate discrepancies in the impact of services and to point to areas of improvement.

Second, from a theoretical perspective, these results with a minimally adapted instrument based on SERVQUAL promote PZB's assertion about the universality of the scale (Parasuraman et al., 1988). This evidence adds another domain to the already proven ones, for which the SERVQUAL scale has revealed to be an adequate instrument with which to assess the quality of service. It also supports the notion that SERVQUAL is adequate to assess all types of services, simple and complex, and that its structure, as well as psychometric properties, are not affected by the nature of the service. It also raises the hypothesis that the adverse findings found in some previous studies could be partly attributed to instruments that displayed excessive departures from the seminal tool (e.g., Carman, 1990).

Third, these results concur with PZB's conceptual and empirically verified five-dimension pattern. The analysis has revealed that not only the five-factor pattern is pertinent, but also that these factors display high level of inter-correlation, indicative of the higher order construct. Accommodating for error in measurement and enabling the specification of the hypothesized relationships between the model's components, confirmatory factor analysis has enabled the fitting of the more appropriate model. Perhaps previous studies, using other methodologies that did not possess the distinctive attributes of *CFA*, uncovered only partially the complex and subtle relationships existing between the measures. Sometimes a single factor would emerge, the global quality construct, while other times a multidimensional, often five-dimension, pattern would transpire. As this study demonstrated, both patterns might have been concurrently justified, but linked hierarchically covert to these researchers.

Fourth, from a methodological perspective, although the simultaneous model will eventually be replaced by the perception-only model on the basis of lack of parsimony, these findings constitute a viable and preferred alternative to using the deficient gap scores. Using confirmatory factor analysis has enabled the simultaneous use of both batteries and the benefit of this rich information, while following rigorous methodological prescriptions.

Research Question 2

The second research question concerns the ability of the data to support equal or different relative importance levels for the quality dimensions. To assess this issue, constrained versions of the perception-only model were estimated and assessed. Fit indices of the alternate models are displayed in Table 4.9.

Setting the five quality structural parameters equal to one constitutes a test of the equal loading of the first-order factors to the second-order quality construct. This model (model 12) was rejected because it failed the likelihood ratio incremental test. Hence, the hypothesis that the dimensions have the same relative weight cannot be confirmed.

In the unconstrained--perception-only-- model, when compared with Tangibles, the other four dimensions carry more similar weights. In the perspective that these four weights may be equal, the difference in the rank-order may not be of statistical or practical significance. To test the hypothesis that these weights are equal, a modified perception-only model (model 13) constraining the second-order factor loading estimates of Reliability, Responsiveness, Assurance, and Empathy to be equal within the battery was tested. Estimation of this model resulted in a degradation of fit indices and this modified model was rejected on the basis of a significant likelihood ratio test.

In the unconstrained model, Responsiveness is the second most important dimension. With the same motivation that drove the previous test, a further test was performed, now focussing on the significance of the rank difference between the most important dimension Assurance, when using the model's implied weights as importance scores, and its runner up, Responsiveness. To assess the hypothesis of equal weights between these two dimensions, an again modified perception-only model was estimated with the parameters for the second-order factor loading estimates of those two dimensions set equal within the dimension (model 14). This hypothesis was rejected at the $p < .05$ level.

Table 4.9.

Fit Indices for the Relative Importance Assessment Models

Model	\bar{T}	df	GFI	RMSEA*	CFI*	NC*	PCFI*	$\Delta \bar{T}$	Δ df
12. Five equal dimensions model	842.92	322	.758	.06	.91	2.62	.83		
Difference between Model 12 and Model 9								115.80 ^{††}	4
13. Four equal dimensions model	748.07	321	.78	.06	.93	2.33	.85		
Difference between Model 13 and Model 9								20.95 ^{††}	3
14. Two equal dimensions model	731.73	319	.78	.06	.93	2.29	.84		
Difference between Model 14 and Model 9								4.62 [†]	1

Note. \bar{T} = Satorra-Bentler robust T statistic; GFI = goodness-of-fit statistic; RMSEA* = robust root mean square error of approximation; CFI* = robust comparative fit index; NC* = robust normed chi-square index; PCFI* = parsimonious comparative fit index; $\Delta \bar{T}$ = likelihood ratio test.

[†] $p < .05$, one-tailed; ^{††} $p < .01$, one tailed.

The weights of the dimensions in the unconstrained perception-only model reveal that the rank-order of the dimensions, when measured as the standardized loading estimates of the dimensions, is different from the relative importance scores. Reliability is not the most important dimension, as in the relative importance scores, but Assurance is. The apparent paradox of the rank-order of the importance of the dimensions needs to be examined.

To investigate this paradox, both the expectations and the perceptions batteries—the simultaneous model—were utilized to enable *ex ante* and *ex post* comparisons. When the standardized higher order factor loading coefficients of both quality batteries from the estimation of the simultaneous model are used as proxies of the relative importance of each dimension in shaping their second-order construct, their rank ordering is not in unison between the batteries. As is evidenced in Table 4.10., the expectations battery displays a

rank-order that is similar to the univariate results. The perceptions battery, however, adopts a different order, with Assurance taking up the lead.

There appears to be unanimity in that Tangibles play a lesser role than the other dimensions in shaping the top-level constructs. The most conspicuous difference between the batteries concerns Reliability, which is top in the empirical measures and expectations loading estimates, but ranks second to last when derived from perceptions. This could be interpreted to mean that respondents externally express their preferences and shape their expectations coherently, but that the shaping of their perceptions adopts a different set of weights.

Table 4.10.

Comparison of the Rank-Order of the Relative Importance of the Dimensions

Dimension	PZB ^a	This Study		
		Relative Importance Items	Expectations Battery Standardized Parameter Estimates	Perceptions Battery Standardized Parameter Estimates
Tangibles	5	5	5	5
Reliability	1	1	1	4
Responsiveness	2	2	2	2
Assurance	3	3	4	1
Empathy	4	4	3	3

^aZeithaml et al., 1990, p. 28.

This observation takes up all its meaning when it is linked to the previous finding that perceptions alone are a good predictor of satisfaction. If the cognitive preferences of residents were not their *true* weights, this could cause a problem to deciders who use these expressed preferences for decision purpose. With the shaping of the quality second-order factor based on a covert set of weights, only the data induced weights are necessary, and the expressed weights redundant.

This finding also partially disputes Carlson and Schwarz' assertion that "factors that determine the quality of service may be weighted differently in the public and private sectors." (1995, p. 27), because it appears that residents cognitively attribute an equal set of weights to commercial and local services. However, it is not known if the weights for their perceptions behave identically in both environments.

The positions of Assurance and Responsiveness as the two most important dimensions for the perceived quality of municipal services make sense for municipal services. Several of the municipal services are provided on a need basis, often on a priority or even an emergency basis. Fire, police and snow removal would typify such services. From the resident's perspective, it is of the utmost importance that the services be provided briskly, even if their performance affects negatively the reliable discharge of some other programmed, but less consequential service activity. A fire or burglar alarm must receive immediate attention, even if it means interrupting some on-going prevention activity.

The success of such interventions is defined more in the outcome than in the conformance to standards. This can be illustrated by an example that occurred in the City of Montréal during the winter of 1997. At that time, the City implemented an array of cost cutting measures that included not removing the snow during weekends. A victim of circumstances, the City suffered repeated heavy snowfalls that made circulation hazardous. Popular discontent grew rapidly and the City resumed snow removal during weekends to improve the situation, while other economy measures appeared to be sustained.

Furthermore, the dominance of Assurance and Responsiveness over Reliability for municipal services can be explained. Residents prefer reliable services, but are ready to subordinate Reliability to more important considerations when circumstances are justified. For instance, residents appreciate a steady, consistent supply of water from the tap, but would view a service interruption as acceptable if required to fight a fire.

In summary, the answer to the second research question is that the data support the assumption of different relative importance levels for the dimensions and that the magnitude

of this importance is best derived from the data. Here again, the results have sizeable methodological implications. First, although dimensions vary in importance, it is not necessary to include items for that purpose in surveys because these weights can be derived directly from the data. Second, it may even be preferable not to collect information from residents on this topic because the evoked set of weights that they will indicate in replies to items may differ from their true set of weights as derived from their perceptions of quality. In addition, not including these items will make the questionnaire even more parsimonious and appealing to respondents, intensifying the previously mentioned benefits. Last, collecting such data may distract users of the information away from the real issues.

For the city managers and elected officials, these results emphasize providing the right service when required rather than focussing on developing uniform response mechanisms. It also challenges the service providers to be attentive to needs rather than concentrating on a mechanistic, production-type quality.

Research Question 3

The third research question was concerned with the ability of the data to support the assumption that the satisfaction scale is valid and reliable for the purpose of representing residents' satisfaction with municipal services.

The validity and reliability assessment performed on the satisfaction scale demonstrated its psychometric properties. The model fit is also good. Consequently, the answer to the third research question is that the data supports the assumption that the satisfaction scale is a valid and reliable model to represent the satisfaction with municipal services.

Research Question 4

The fourth research question is concerned with the ability of the data to support either the simultaneous model or the perception-only model to represent the relationship between quality and satisfaction with municipal services. Although the hypothesis that the

simultaneous model is adequate to represent the relationship between the variable in the data set has not been rejected, the more parsimonious perception-only model has been retained. The validity and reliability assessment indicated that the perception-only model scales' psychometric properties were demonstrated, except for slight weaknesses. Consequently, the answer to the fourth research question is that the perception-only scale best represents the relationship between quality and satisfaction with municipal services.

This finding falls in line with the stream of authors who found evidence in favour of the perception-only predictive instrument (e.g., Cronin & Taylor, 1992), a conclusion that has been endorsed by the authors of the scale (Parasuraman et al., 1993). It also has several methodological and managerial implications. First, the now simplified tool is a valuable instrument to assess the perceived quality of service. The five-dimension quality pattern is confirmed and the universality argument is not affected. Hence expectations are not undesirable, they have merely become redundant.

Second, the property of the instrument to capture attitudes concerning the outcome of the service has also been demonstrated. Not only is the instrument pertinent to the assessment of the quality of municipal services, but it has also proven to be a valid and reliable tool to gauge the satisfaction towards these services, as well as a suitable mechanism to link them causally.

Third, the presence of expectations does not seem to improve the predicting power of the instrument. Thus, the utility of this battery is considerably reduced and it is possibly superfluous. If data about expectations need not be collected, this resolves the dilemma concerning the usefulness of measuring expectations and the manner in which such data collection should be performed.

Managerially, it demonstrates several advantages. First, the reduced instrument constitutes a parsimonious tool, which administration might be preferable to the expanded instrument. Its brevity will translate in reduced production as well as distribution costs. Shorter, it will be more appealing to the respondents and it should foster better response

rates as well as fewer incomplete cases. Finally, the presence of a single battery will alleviate any possible confusion in the respondents' minds caused by two similarly worded sets of items.

Second, the single perspective to quality will focus managers and first-line employees alike towards *the* factor of influence. In that perspective, considering both expectations as well as perceptions may constitute a distraction and dilute the problems that poor quality levels may indicate. Having a single lever to manipulate should increase the concentration and dedication towards perceptual quality improvements.

Investigating Factorial Invariance

Also of interest is the tenability of the model in sub-sets of the population. To assess the proposition that the model would fit equally well in all segments of the population, multiple group tests were performed with the data set partitioned in homogeneous groups determined from the socio-demographic attributes of the population. For simplicity, the groups formed here were the same as those used when assessing the representativeness of the sample. The multiple group tests assessed the factorial invariance of the measurement and structural parameters--the equality of the relationships between the indicators and the factors as well as between the factors. Due to software limitations, the \bar{T} and robust statistics are not available when performing this procedure. Thus, the results of the tests were obtained from the likelihood ratio test based on the T statistic and the Lagrangian Multiplier tests on the across-group equality parameters. For such tests, significant values are indicative that the parameters were not equal in the sub-samples. For the likelihood ratio test, at $\alpha = .05$, a T statistic difference of 38.89 or less between the constrained and the unconstrained models ($df=26$) is indicative of equivalent fit. Because the T statistic is biased, some of the tests results may not hold should the unbiased \bar{T} statistic been available. Table 4.11. displays the results of the estimation of the constrained and unconstrained models.

Table 4.11.
Statistics for the Multigroup Analysis

Attribute and Model	<i>T</i>	ΔT
Age (18-44 years, 45 years or more)		
1. Unconstrained model	1834.83	
2. Partially constrained model	1855.46	
3. Constrained model	1886.48	
Difference between Model 2 and Model 1		20.63
Difference between Model 3 and Model 1		51.65 [†]
Education		
1. Unconstrained model	1874.59	
2. Partially constrained model	1892.41	
3. Constrained model	1915.48	
Difference between Model 2 and Model 1		17.82
Difference between Model 3 and Model 1		40.89 [†]
Employment ^a		
1. Unconstrained model	--	
2. Partially constrained model	--	
3. Constrained model	--	
Difference between Model 2 and Model 1	--	
Difference between Model 3 and Model 1	--	
Gender		
1. Unconstrained model	1832.39	
2. Partially constrained model	--	
3. Constrained model	1863.36	
Difference between Model 2 and Model 1		--
Difference between Model 3 and Model 1		30.97
Income		
1. Unconstrained model	1875.38	
2. Partially constrained model	--	
3. Constrained model	1900.62	
Difference between Model 2 and Model 1		--
Difference between Model 3 and Model 1		25.24
Ownership		
1. Unconstrained model	1859.90	
2. Partially constrained model	1893.12	
3. Constrained model	1916.50	
Difference between Model 2 and Model 1		33.22
Difference between Model 3 and Model 1		56.60 [†]
Stability		
1. Unconstrained model	1774.33	
2. Partially constrained model	1792.92	
3. Constrained model	1826.43	
Difference between Model 2 and Model 1		18.59
Difference between Model 3 and Model 1		52.10 [†]

Note. *Df* for the unconstrained model = 636; *df* for the partially constrained model = 658; *df* for the constrained model = 662; *df* for the difference between Model 2 and Model 1 = 22; *df* for the difference between Model 3 and Model 1 = 26. Dashes (--) indicate that no value was obtained for this model or test. $T = \chi^2$ distributed estimator of model fit.

^aA proper solution was not obtained for the multigroup analysis on the employment variable.

[†] $p < .05$, one-tailed.

For men and women, the factorial invariance is tenable on all parameters, indicating that there is no difference between the genders in the shaping of the quality and satisfaction relationships. This result is also observed for the partitioned income groups. Attempts to test in subgroups the difference of the employment variable failed to yield a proper solution and no other practical grouping was available. For the other three attributes, some differences between the groups were observed.

For education, the partitioning of residents with high school education or less against residents who attended CEGEP and university indicated that the equality of parameters assumption did not hold. Four indicator-to-factor links were found delinquent as manifested through the LM test. At the same significance level, the estimation of a partially constrained model releasing these four equality constraints revealed a fit equivalent to the unconstrained model ($T \leq 33.92$, $df = 22$). The ownership of residence criteria displayed a similar pattern and a partially constrained model resolved the four delinquent indicator-to-factor links. For stability of residence, one of four problematic links joined Empathy to the second-order factor and the other three involved indicators and factors; this was also resolved successfully by a partially constrained model that freed the problematic links.

Attempts to perform the same test on age subgroups initially failed to deliver a proper solution. This was attributed to computational difficulties encountered in view of the asymmetric sample sizes. When age subgroups were clustered at the 45 years threshold, the sample sizes became more alike and a proper solution was obtained. A significant difference was observed. Three items to factor links were delinquent, as well as the Reliability to second-order factor link. A partially constrained model releasing the problematic constraints was found equivalent to the unconstrained model.

In conclusion, the results of the analysis indicate that the instrument describes equally well the perceptions of residents of both genders and different income levels. There appears to be some differences on the basis of the residents' education, ownership of

residence, and stability of residence. However, these differences seem mild and may lack practical significance. These finding add value to the instrument in that this homogeneity reinforces the universal character of the scales.

SUMMARY AND CONCLUSION

In this section, the findings will be reviewed first. Then, the implications of these findings will be examined. Next, the limitations of the study will be reviewed and future possible research tracks will be outlined. In conclusion, the contributions of this research will be outlined.

Summary of the Findings

Because univariate measures are prone to error in measurement, they were not used in the current study for the formal investigation of research questions, but they are useful when compared to similar results obtained in previous studies. One such finding is that residents do not set their desires at the utopian level. In the current study, the measures of the expectations scale were well below the scale maximum. This indicates that the expectations of residents towards the services they receive are realistic. However, a subsequent finding was that expectations are not very important in the shaping of satisfaction, making the level of expectations less relevant.

The first specific aspect of this research was concerned with the validity and the reliability of the second-order factor model to measure the perceived quality of municipal services. The model fitted well and the psychometric properties of the instrument were ascertained. Thus, its utility is confirmed. In addition, a framework to reconcile these findings with previous evidence was elaborated.

The second research question investigated the equality or inequality of the relative importance of the dimensions in shaping the higher order quality construct. The hypothesis that the dimensions carry equal weight was not retained. In addition, a paradox was discovered between the importance of the dimensions as expressed cognitively and those

that were model implied, because the set of weights for expectations appears to differ from that for perceptions.

The third research question was devoted to verifying the validity and the reliability of the instrument to measure the satisfaction of residents towards municipal services. The model fitted well and the psychometric properties of these scales were also verified. Thus, the utility of the instrument was confirmed.

The fourth research question centred on the appropriateness of either the simultaneous model or the perception-only model to express the relationship between perceived quality and satisfaction. The simultaneous model was found adequate to represent this relationship, but the performance-based model was preferred because it achieves the same aim in a more parsimony fashion. A subsidiary test for the factorial invariance of the instrument across homogeneous sub-samples was performed. The result is that the socio-demographic attributes of the respondents may affect the factor pattern of the instrument, but in no material way.

Implications

Through the investigation of these questions, the current research has made several findings that enrich the pertinent body of knowledge. In particular, this study has addressed the issues raised about the SERVQUAL scale and its findings shed light on most issues, either confirming previous findings or hypothesizing a new answer. The pertinent findings are discussed next. The implications are grouped as (a) practical, (b) theoretical, or (c) methodological.

Practical Implications

First, for municipal deciders who want to externally assess the quality of the services that they deliver to their constituents, the perception-only model scales constitute a valuable and practical tool. The quality and satisfaction scales' psychometric properties

have been established. This will enable city officials to supplement the internal measures of quality with the attitudes of the residents towards these services. The heuristic measures obtained through this instrument can be used to corroborate factual information. In addition, they bring a fresh perspective when compared to operational measures that cater to a more mechanistic notion of quality. The instrument can be used to confirm weaknesses in the service uncovered in the traditional manner and to ferret other areas of improvement that were not uncovered by the process indicators.

Second, a significant causal link was found between quality of municipal services and satisfaction of the residents with these services, with higher quality services leading to superior levels of satisfaction. This fresh evidence reveals the importance of high quality services to satisfied residents. From a managerial perspective, this finding constitutes a clear beacon on which to focus when selecting between alternate courses of action in service delivery. The service endeavours most likely to impact positively on the satisfaction of residents are those that possess higher levels of perceived quality, as viewed from the residents' perspective.

Third, parsimony constitutes a significant feature of the instrument. The psychometric properties of the instrument have been demonstrated both for the quality and satisfaction scales of the instrument. Because the expectation battery is not an important factor in the shaping of the attitude towards satisfaction and because the relative importance scale is redundant, the instrument adopts a reduced form. There are several practical advantages to a compressed instrument. From a logistic perspective, a shorter questionnaire will result in important savings in production and distributions costs. In terms of responses, its brevity should induce (a) more replies, because the time cost of completing the survey will decrease; (b) more complete replies, because fewer respondents will get tired or discouraged before the end, reducing the reliance on imputed values; and (c) better replies, because the possible confusion incurred by some respondents when faced with two sets of items with a similar wording has been alleviated.

Fourth, the instrument developed in the course of this study stems from a reputable tool, SERVQUAL, which has been used in a considerable number of settings. Only a minimal adaptation was imposed to the SERVQUAL scale in the development of the current instrument. Thus, the worth of the current instrument is considerably enhanced by its close conceptual and operational proximity to the source. This should reduce considerably the perceived psychological risk of adopting an unproven methodology. To the contrary, the current instrument benefits from all the previous empirical tests to which SERVQUAL was subjected. Public deciders may consider adopting the current instrument not only on its intrinsic value but also on the reputation warranted by its *pedigree*.

Fifth, focussing on perceptions alone measured through the internally derived importance of the dimensions will orient all levels of municipal staff towards an error-free single set of priorities. This single yardstick will concentrate the attention of all parties on the perceptions, which alone influence the satisfaction of residents. In the expectations/perceptions disconfirmation paradigm, it is possible to contemplate improvements in quality by expectations management alone. This approach may be viewed as opportunistic and not likely to produce sustained effects. Rather, focussing on *the* single lever is bound to promote plans of action with longer-term effects.

Theoretical Implications

First, the current research has found that a SERVQUAL derived instrument proved to be a valid and reliable instrument to assess the perceived quality of municipal services. The five dimensions of service quality have proven their usefulness in the assessment of local services. The municipal services domain thus constitutes an additional sector to which the utility of the instrument has been extended. Hence, this study constitutes evidence in favour of the universality of the SERVQUAL scale.

Second, the results concur generally with PZB's theoretical framework of quality. The empirical evidence supports the idea that quality encompasses five dimensions, which collectively define the domain. The use of *CFA* has enabled the estimation of a well fitting

simple structure that embodies the conceptual model. As expected, the high inter-correlation values between the dimensions testify to their close proximity.

However, the findings of this study diverge from PZB's work in that the quality construct is not modelled through correlated first-order factors, but as a second-order factor model. A model linking the quality dimensions to a second-order factor was successfully estimated and its fit was found appropriate. The current study thus confirms the assumption that the dimensions of SERVQUAL can be joined in a hierarchical manner.

The repeated claim that SERVQUAL items load on a single factor has also been addressed. The *EFA* results in this study would also tend to support the unidimensionality claim. However, the single factor model hypothesis was not supported when estimated through *CFA*, because a poor fit was observed. Thus, the conclusion should not be that the scale is unidimensional. The conclusion should be that, when the *EFA* procedure is used, the strong link between all items sometimes takes precedence over the five-prong articulated structure and reveals a single factor. At other times when the *EFA* procedure is used, the top factor is present but not as prevalent, and a five-factor pattern is extracted instead. As the first/second-order factor relationships between the variables overlap in all data sets, the emergence of one or the other solution is merely a sample effect.

This also points to the more precise specification properties of *CFA*. These features become distinctly advantageous over *EFA* as the relationships between the variables get intensely complex, especially because *EFA* cannot accommodate for hierarchical ramifications.

Third, the current study does not support the disconfirmation paradigm approach to defining quality. The comparison of perceptions to expectations did not appear to be material in the shaping of satisfaction. To the contrary, perceptions alone proved to be equally effective, as measured by the incremental improvement in fit when passing from the theoretical to the reduced model. These findings confirm the studies that found that perceptions alone influence satisfaction. Of particular interest is the fact that the factor

pattern of the perceptions and satisfaction constructs was preserved despite the simplification of the instrument. More generally, the results of this study support the previous evidence on perceptual measures as significant predictors of satisfaction.

Fourth, factorial invariance tests on the instrument have indicated that there is no interaction between gender or income and the results of the instrument. While age, education, ownership, and stability of residence have a slight influence on the factor pattern of the instrument, the differences between the groups are small and may not have any practical effects.

Fifth, accounting for error through the use of *CFA* has shed light on the complex pattern of the relative importance of the quality dimensions in shaping the higher order construct. *True* importance levels were derived from the loading estimates of the dimensions on their respective second-order constructs. The equal importance hypothesis has been rejected.

The study has also shown that residents, perhaps unconsciously, attribute a different set of weights to the shaping of their perceptions from the weights that they confer to their expectations. The weights used for the expectations towards local services are aligned with the expressed relative importance scores and are similar to those reported in previous studies towards commercial services. The weights for perceptions, however, are rank-ordered differently and express some other set of preferences. Reliability has been advocated as the most important dimension for commercial services, but ranks lower in connection with the perceptions towards local services. For these measures, Assurance, followed by Responsiveness, takes the lead. The salience of these dimensions can be justified for local services by the priority manner in which several of these services are delivered. When a service such as police or snow removal is performed, there is a greater concern by the resident recipient for an energetic delivery focussed on effectiveness, even to the detriment of the smooth, routine delivery of some other, less consequential service.

Methodological Implications

First, the quality resources of the instrument are minimally different from SERVQUAL. The results of this study generally support the five-dimension pattern of SERVQUAL and some of the empirical results obtained by PZB. This raises the issue of the adaptation of the scale to particular settings. There is no dispute that any adaptation that may raise the usefulness of an instrument should be undertaken and that any item that will provide useful information on the matter being investigated should be included in the survey. However, the more different the revised instrument is from SERVQUAL, the less pertinent any comparison with the original becomes. Perhaps some of the differences from SERVQUAL stemming from instruments that were reduced to 15 or augmented to 40 items were attributable in part to these modifications.

Secondly, the approach of computing difference scores as indicators of quality, which is not considered a sound practice by many researchers, was circumvented successfully in this study. The challenge of accounting for the influence of both expectations and perceptions measures on the shaping of satisfaction was achieved through the use of *SEM*. The use of this flexible methodology has enabled the simultaneous incorporation of both groups of attitudes in a model and fostered the assessment of their respective contributions. However, a reduced model was found more pertinent and parsimonious, reducing the utility of the simultaneous model. Nonetheless, the simultaneous model was successful in providing an innovative and rigorous alternative to the gap scores.

Third, the issue of contemporaneous measurement of perceptions and expectations was resolved, albeit indirectly. The problem with measuring expectations and perceptions with the same instrument was addressed by setting the expectations at the ideal level, making them invariant in time. Better still, the little influence of expectations on satisfaction does not justify that these items be maintained in the survey questionnaire, settling the war for lack of warriors.

Limitations

The results of this study are subject to some limitations that may reduce their external validity. First, the restricted number of cases available for analysis has caused computational difficulties and a larger sample might have yielded a solution for all analyses. In addition, the small sample size may induce that these results will not be taken seriously, because "the usual low response rates [stemming from the use of mail surveys], typically 25% or considerably less, prevent such findings from being credible" (Hatry & Blair, 1976, p. 135). This problem had been anticipated and many steps had been taken to encourage respondents to complete the questionnaire. It is possible that a follow-up note sent to all sampled units would have had a stronger effect on non-respondents than the reminder via the media.

Second, the results may be considered representative exclusively for the survey population—all the residents of the Town of Saint-Jean. They cannot be extended directly to residents of other towns. There may be several factors that have not been controlled for and that could influence the results, such as the size of the town, the geographic location, etc. More data must be collected from a representative random sample of the population of other towns to assess if the results found are equally valuable for explaining satisfaction with municipal services in those settings.

Third, this study focussed on the relationship between perceived quality and satisfaction with municipal services. Perhaps other factors, which were not considered in this study, may influence satisfaction with the town's services. For instance, the effect of taxation may not be neutral and trends in satisfaction may be displayed for different levels of tax disbursements. To establish perceived quality as the *sole* motivator of residents' satisfaction, all these potential factors must be conceptualized, empirically tested, and ruled out. Consequently, the results of this study concern a good fitting probable model and do not undermine the hypothesis that other models, with different patterns, may fit the data better.

Prospects for the Future

Although this study has found suitable replies to some questions concerning the quality and satisfaction with municipal services, it has also raised several new questions worth investigating.

First, as mentioned in the previous sub-section, the scope of this research was limited. While a significant relationship has been evidenced, an unknown number of factors have not been tested. The satisfaction of residents has been an object of academic research for some time, but the factors influencing satisfaction are still largely undetermined. This study has highlighted only one of perhaps several factors that may play a significant role in the satisfaction of residents. Because this topic has recently become of greater interest to public officials, more efforts will likely be engaged to solve this issue. Further studies could accommodate for factors such as the tax disbursements of residents and attempt to link them significantly to satisfaction. This is required to ascertain the shaping of satisfaction with municipal services, in order to target the mechanisms most likely to improve it.

Second, and also raised in the previous sub-section, these results cannot be generalized immediately. While it is interesting to have these results for Saint-Jean, it would be more useful if similar findings could be obtained for a large group of towns, such as all the towns of the Province of Québec. The residents of more towns need to be surveyed to obtain common results. The study could be pursued efficiently, through randomness in the sampling scheme, for optimal representativeness. Such a study would serve to confirm or infirm the preliminary results provided by the current study, potentially controlling for discriminant factors such as the size of town, etc. This line of investigation would validate whether or not the instrument is fit for general use with an extended population.

Third, this study has examined the phenomenon from a static, reactive point-of-view. Of considerable managerial interest would be the adoption of a dynamic, pro-active

perspective. This would entail investigating organizational and operational practices within town structures that most likely have a positive effect on satisfaction. This would parallel similar studies that have identified some of these factors for the commercial sector (e.g., Waldman & Gopalakrishnan, 1996). This line of investigation would have considerable importance for town officials in that it may result in defining sets of best practices that foster quality and satisfying services in the municipal sector. Using these context specific measures would advantageously replace the current practice of blindly mimicking the private sector initiatives, which is used with the untested assumption that the results provided are as beneficial in the municipal as in the commercial context.

Fourth, the hypothesis that satisfaction and other behavioural intentions of the residents coalesce has yet to be verified. Besides satisfaction, the perceived quality of municipal services may, or may not, have other effects on residents such as encouraging renovation of real estate. Testing for these potential relationships could provide valuable knowledge to elected officials on actions to undertake in order to influence such behaviours in a desirable way.

Fifth, the findings of this study concerning SERVQUAL raise new issues about the scale. For example, the dual set of importance weights that was identified in this study might also prevail in other environments. To investigate this, new data need not be collected because data sets that served as the basis for previous studies can be re-examined using the methodology elaborated here.

Contributions

In conclusion, this study has made several contributions to the extant body of knowledge. These contributions are centred on the properties of the SERVQUAL scale and on the measurement of perceived quality and satisfaction with municipal services.

First, this study has successfully addressed several of the outstanding issues concerning the SERVQUAL scale. In addition, this study has raised and partially solved an

additional issue with the scale. Second, this study has contributed to the development of a valid, reliable, and parsimonious instrument to assess quality and satisfaction with municipal services. Should the utility of the instrument be extended to a larger population, the use of this tool will yield considerable benefits to municipal decision-makers, and by a ripple effect, to residents.

In summary, this study has demonstrated the utility of the SERVQUAL scale when used towards municipal services. It has evidenced the critical role of perceptual measures of quality to the shaping of satisfaction with local services. Finally, it has hinted at some opportunities for further investigations concerning the quality of municipal services.

REFERENCES

- Ahire, S. L., Golhar, D. Y., & Waller, M. A. (1996). Development and validation of TQM implementation constructs. *Decision Sciences*, 27(1), 23-56.
- Alreck, P. L., & Settle, R. B. (1995). *Survey research handbook*. (2nd ed.). Chicago: Irwin.
- Alwin, D. F., & Krosnick, J. A. (1991). The reliability of survey attitude measurement. *Sociological Methods & Research*, 20(1), 139-181.
- Anderson, J. C., & Gerbing, D. W. (1984). The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika*, 49(2), 155-173.
- Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*, 103(3), 411-423.
- Andreassen, T. W. (1994). Satisfaction, loyalty and reputation as indicators of customer orientation in the public sector. *International Journal of Public Sector Management*, 7(2), 16-34.
- Armstrong, J. S., & Overton, T. S. (1977). Estimating nonresponse bias in mail surveys. *Journal of Marketing Research*, 14(August), 396-402.
- Asubonteng, P., McLeary, K. J., & Swan, J. E. (1996). SERVQUAL revisited: A critical review of service quality. *Journal of Services Marketing*, 10(6), 62-81.
- Babakus, E., & Boller, G. W. (1992). An empirical assessment of the SERVQUAL scale. *Journal of Business Research*, 24, 253-268.
- Babakus, E., & Mangold, W. G. (1989). Adapting the 'SERVQUAL' scale to health care environment: An empirical assessment. In P. Bloom, R. Winer, H. H. Kassarian, D. L. Scammon, B. Weitz, R. E. Spekman, et al. (Eds.), *Enhancing knowledge development in marketing: AMA educators' proceedings*, (Vol. 1, p. 195). Chicago: American Marketing Association.
- Babakus, E., Pedrick, D. L., & Inhofe, M. (1993). Empirical examination of a direct measure of perceived service quality using 'SERVQUAL' items. In D. W. Cravens & P. R. Dickson (Eds.), *Enhancing knowledge development in marketing: AMA educators' proceedings*, (Vol. 4, pp. 8-9). Chicago: American Marketing Association.
- Bagozzi, R. P., & Heatherton, T. F. (1994). A general approach to representing multifaceted personality constructs: Application to state self-esteem. *Structural Equation Modeling*, 1(1), 35-67.
- Barnett, V., & Lewis, T. (1994). *Outliers in statistical data*. (3rd ed.). Chichester, UK: Wiley.

Bartlett, M. S. (1954). A note on the multiplying of factors for chi-square approximations. *Journal of the Royal Statistical Society*, 16(2), 296-298.

Battle, B. F., & Nayak, P. R. (1994). Going public. *Spectrum: The Journal of State Government*, 67(2), 16-22.

Baumgartner, H., & Homburg, C. (1996). Applications of structural equation modeling in marketing and consumer research: A review. *International Journal of Research in Marketing*, 13, 139-161.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107(2), 238-246.

Bentler, P. M. (1995). *EQS Structural equations program manual*. Encino, CA: Multivariate Software.

Bentler, P. M., & Chou, C.-P. (1988). Practical issues in structural modeling. In J. S. Long (Ed.), *Common problems/Proper solutions: Avoiding error in quantitative research*, (pp. 161-192). Newbury Park, CA: Sage.

Bentler, P. M., & Dudgeon, P. (1996). Covariance structure analysis: Statistical procedure, theory, and directions. In J. T. Spence, J. M. Darley, & D. J. Foss (Eds.), *Annual Review of Psychology*, (Vol. 47, pp. 541-570). Palo Alto, CA: Annual Reviews.

Bentler, P. M., & Wu, E. J. C. (1996). *EQS for Power Macintosh 5.4* (Release Notes). Encino, CA: Multivariate Software.

Berry, L. L., Parasuraman, A., & Zeithaml, V. (1988). The service quality puzzle. *Business Horizons*, 31(September-October), 35-43.

Berry, L. L., Parasuraman, A., & Zeithaml, V. A. (1994). Improving service quality in America: Lessons learned. *Harvard Business Review*, 72(November-December), 101-112.

Bishop, G. F., Oldenick, R. W., Tuchfarber, A. J., & Bennett, S. E. (1980). Pseudo-opinions on public affairs. *Public Opinion Quarterly*, 44(2), 198-209.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: John Wiley & Sons.

Bollen, K. A., & Long, J. S. (Eds.). (1993). *Testing structural equation models*. Newbury Park, CA: Sage.

Boulding, W., Kalra, A., Staelin, R., & Zeithaml, V. A. (1993). A dynamic process model of service quality: From expectations to behavioral intentions. *Journal of Marketing Research*, 30(February), 7-27.

Bovaird, T. (1996). Performance assessment of service quality: Lessons from UK national initiatives to influence local government. In H. Hill, H. Klages, & E. Löffler (Eds.), *Quality, innovation and measurement in the public sector*, (pp. 37-65). Frankfurt, Germany: Peter Lang.

Boyt, T. E. (1994). *A study of consumer perception of quality based on price, outcome, and information*. Unpublished doctoral dissertation, University of Oklahoma, Norman.

Brede, L., Jr. (1985). *The correlates of soldier satisfaction with police services in army communities*. Unpublished doctoral dissertation, University of Georgia, Athens.

Brown, J. R., & Fern, E. F. (1981). Goods vs services marketing: A divergent perspective. In J. H. Donnelly & W. R. George (Eds.), *Marketing of services*, (pp. 205-207). Chicago: American Marketing Association.

Brown, K., & Coulter, P. B. (1983). Subjective and objective measures of police service delivery. *Public Administration Review*, 43(January/February), 50-59.

Brown, S. (1990, May 28). Don't leave 'customers' out of the health care quality dialogue. *Marketing News*, 24, 14-15.

Brown, T. J., Churchill, G. A., Jr., & Peter, J. P. (1993). Improving the measurement of service quality. *Journal of Retailing*, 69(1), 127-139.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models*, (pp. 136-162). Newbury Park, CA: Sage.

Brudney, J. L., & England, R. E. (1982). Urban policy making and subjective service evaluations: Are they compatible? *Public Administration Review*, 42(March/April), 127-135.

Buckley, P. J., Pass, C. L., & Prescott, K. (1992). The internationalization of service firms: A comparison with the manufacturing sector. *Scandinavian International Business Review*, 1(1), 39-56.

Byrne, B. M. (1994). *Structural equation modeling with EQS and EQS/Windows: Basic concepts, applications and programming*. Thousand Oaks, CA: Sage.

Campbell, B. D. (1994). *Members' perception of service quality at the Nellis Air Force Base Officers' Open Mess*. Unpublished master thesis, University of Nevada, Las Vegas.

Campbell, D. T. (1960). Recommendations for APA tests standards regarding construct, trait, or discriminant validity. *American Psychologist*, 15(August), 546-553.

Canada Post Corporation. (1996). *Householder figures for urban delivery installations - Quebec*. Ottawa, Canada: Canada Post.

Carlson, M. S., & Schwarz, R. M. (1995). What do citizens really want? Developing a public-sector model of service quality. *Popular Government*, 60(Spring), 26-33.

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

Carmines, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment*. (Vol. 07-017). Newbury Park, CA: Sage.

- Carroll, S. (1978). *An analysis of the relationship between citizen perceptions and unobtrusive measures of street conditions* (Research Report, M-10). Bloomington: Indiana University, Workshop in Political Theory and Public Policy.
- Cattell, R. B. (1956). Validation and intensification of the sixteen personality factor questionnaire. *Journal of Clinical Psychology*, 12, 205-214.
- Cattell, R. B. (1966). The meaning and strategic use of factor analysis. In R. B. Cattell (Ed.), *Handbook of multivariate experimental psychology*, (pp. 174-243). Chicago: Rand McNally.
- CEPPP. (1992). *Considering quality: An analytical guide to the literature on quality and standards in the public services*. Uxbridge, UK: Centre for the Evaluation of Public Policy and Practice, Brunel University.
- Chen, I. J., Gupta, A., & Rom, W. (1994). A study of price and quality in service operations. *International Journal of Service Industry Management*, 5(2), 23-33.
- Cigler, B. A. (1996). Adjusting to changing expectations at the local level. In J. L. Perry (Ed.), *Handbook of public administration*, (2nd ed., pp. 60-76). San Francisco: Jossey-Bass.
- Comrey, A. (1973). *A first course in factor analysis*. New York: Academic Press.
- Creswell, J. W. (1994). *Research design: Qualitative & quantitative approaches*. Thousand Oaks, CA: Sage.
- Crompton, J. L., & Mackay, K. J. (1989). User's perceptions of the relative importance of service quality dimensions in selected public recreation programs. *Leisure Sciences*, 11, 367-375.
- Cronbach, L. J. (1951). Coefficient alpha and internal structure of tests. *Psychometrika*, 16(3), 297-334.
- Cronbach, L. J., & Furby, L. (1970). How we should measure "change" --or should we? *Psychological Bulletin*, 74(1), 68-80.
- Cronin, J. J., Jr., & Taylor, S. A. (1992). Measuring service quality: A reexamination and extension. *Journal of Marketing*, 56(July), 55-68.
- Cronin J. J., Jr., & Taylor, S. A. (1994). SERVPERF versus SERVQUAL: Reconciling performance-based and perceptions-minus-expectations measurement of service quality. *Journal of Marketing*, 58(January), 125-131.
- Crosby, P. B. (1979). *Quality is free*. New York: McGraw-Hill.
- Dabholkar, P. A. (1993). Customer satisfaction and service quality: Two constructs or one? In D. W. Cravens & P. R. Dickson (Eds.), *Enhancing knowledge development in marketing: AMA educators' proceedings*, (Vol. 4, pp. 10-18). Chicago: American Marketing Association.
- Darby, M. R., & Karni, E. (1973). Free competition and the optimal amount of fraud. *Journal of Law and Economics*, 16, 67-86.

Das, H., Das, M., & McKenzie, F. (1995). Assessing the 'will of the people': An investigation into town service delivery satisfaction. *Canadian Public Administration*, 38(1), 77-93.

Davis, D., & Cosenza, R. M. (1993). *Business research for decision making*. (3rd ed.). Belmont, CA: Wadsworth.

Ding, L., Velicer, W. F., & Harlow, L. L. (1995). Effects of estimation methods, number of indicators per factor, and improper solutions on structural equation modeling fit indices. *Structural Equation Modeling*, 2(2), 119-143.

Donnelly, M., Wisniewski, M., Dalrymple, J. F., & Curry, A. C. (1995). Measuring service quality in local government: The SERVQUAL approach. *International Journal of Public Sector Management*, 8(7), 15-20.

Dorfman, P. W. (1979). Measuring and meaning of recreation satisfaction. *Environment and Behavior*, 11(4), 483-510.

Drebin, A., & Brannon, M. (1990). Police department programs. In H. P. Hatry, J. R. Fountain, Jr., J. M. Sullivan, & L. Kremer (Eds.), *Service efforts and accomplishments reporting: Its time has come - An overview*, (Vol. GR07, pp. 308). Norwalk, CT: Government Accounting Standards Board.

Eaton, B. C., & Eaton, D. F. (1995). *Microeconomics*. (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Edwards, J. R. (1994). The study of congruence in organizational behavior research: Critique and proposed alternative. *Organizational Behavior and Human Decision Processes*, 58, 51-100.

Edwards, J. R. (1995). Alternatives to difference scores as dependent variables in the study of congruence in organizational research. *Organizational Behavior and Human Decision Processes*, 64(3), 307-324.

Etezadi-Amoli, J., & Farhoomand, A. F. (1996, 2-4 July). *The end user computing success construct*. Paper presented at the 4th European Conference on Information Systems, Lisbon, Portugal.

Evans, J. R., & Lindsay, W. M. (1996). *The management and control of quality*. (3rd ed.). St. Paul, MN: West.

Finn, D. W., & Lamb, C. W., Jr. (1991). An evaluation of SERVQUAL scales in a retailing setting. In R. H. Holman & M. R. Salomon (Eds.), *Advances in Consumer Research*, (Vol. 18, pp. 483-490). Provo, UT: Association for Consumer Research.

Folz, D. H. (1996). *Survey Research for Public Administration*. Thousand Oaks, CA: Sage.

Ford, J. B., Joseph, M., & Joseph, B. (1993). Service quality in higher education: A comparison of universities in the United States and New Zealand using SERVQUAL. In D. W. Cravens & P. R. Dickson (Eds.), *Enhancing knowledge development in marketing: AMA educators' proceedings*, (Vol. 4, pp. 75-81). Chicago: American Marketing Association.

Galloway, R. (1992, 10-12 February). *Improving the quality of police service using private sector techniques*. Paper presented at the First International Symposium on Productivity and Quality Improvement with a Focus on Government, Washington, DC.

Garvin, D. A. (1983). Quality on the line. *Harvard Business Review*, 61(September/October), 65-73.

Garvin, D. A. (1987). Competing on the eight dimensions of quality. *Harvard Business Review*, 65(November/December), 101-109.

Gerbing, D. W., & Anderson, J. C. (1988). An updated paradigm for scale development incorporating unidimensionality and its assessment. *Journal of Marketing Research*, 25(May), 186-192.

Glauser, M. J., & Tullar, W. L. (1985). Citizen satisfaction with police officer/citizen interaction: Implications for the changing role of police organizations. *Journal of Applied Psychology*, 70(3), 514-527.

Goodsell, C. T. (1981). The contented older client of bureaucracy. *International Journal of Aging and Human Development*, 14(1), 1-9.

Gorey, K. M., Chandler, R. G., & Osmun, D. (1996). Prevalent department of social services' client satisfaction: A research and practice note. *Canadian Journal of Program Evaluation*, 11(1), 129-138.

Gotlieb, J. B., Grewal, D., & Brown, S. W. (1994). Consumer satisfaction and perceived quality: Complementary or divergent constructs? *Journal of Applied Psychology*, 79(6), 875-885.

Grönroos, C. (1983). *Strategic management and marketing in the service sector* (Report 83-104). Cambridge, MA: Marketing Science Institute.

Grönroos, C. (1993). Toward a third phase in service quality research: Challenges and future directions. In T. A. Swartz, D. E. Bowen, & S. W. Brown (Eds.), *Advances in service marketing and management: Research and practice*, (Vol. 2, pp. 49-64). Greenwich, CT: JAI Press.

Hair, J. F., Jr., Anderson, R. E., Tatham, R. L., & Black, W. C. (1992). *Multivariate data analysis: With readings*. (3rd ed.). New York: Macmillan.

Harvey, J., Lefebvre, É., & Lefebvre, L. A. (1993). Technology and the creation of value in services: A conceptual model. *Technovation*, 13(8), 481-495.

Hatry, H. P., & Blair, L. H. (1976). Citizen surveys for local government: A copout, manipulative tool, or a policy guidance and analysis aid? In T. N. Clark (Ed.), *Citizens preferences and urban public policy: Models, measures, uses*, (Vol. 34, pp. 129-141). Beverly Hills, CA: Sage.

Hatry, H. P., Blair, L. H., Fisk, D. M., Greiner, J. H., Hall, J. R., Jr., & Schaenman, P. S. (1977). *How effective are your community services?* Washington, DC: Urban Institute Press and International City Management Association.

Hayduk, L. A. (1996). *LISREL Issues, debates, and strategies*. Baltimore: Johns Hopkins University Press.

Hayes, B. E. (1991). *Measuring customer satisfaction: Development and use of questionnaires*. Milwaukee, WI: American Society for Quality Control Press.

Hu, L.-T., Bentler, P. M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin*, 112(2), 351-362.

Hubbard, R., & Allen, S. J. (1987). A cautionary note on the use of principal components analysis: Supportive empirical evidence. *Sociological Methods & Research*, 16, 301-308.

Iacobucci, D., Ostrom, A., & Grayson, K. (1995). Distinguishing service quality and consumer satisfaction: The voice of the consumer. *Journal of Consumer Psychology*, 4(3), 277-303.

James, L. R., Mulaik, S. A., & Brett, J. M. (1982). *Causal analysis: Assumptions, models, and data*. (Vol. 1). Beverly Hills, CA: Sage.

Johns, G. (1981). Difference score measures of organizational behavior variables: A critique. *Organizational Behavior and Human Performance*, 27, 443-463.

Johnston, R., Silvestro, R., Fitzgerald, L., & Voss, C. (1990, June 05-08). *Developing the determinants of service quality*. Paper presented at the Séminaire International de Recherche en Management des Activités de Service, LaLonde Les Maures, France.

Jöreskog, K. G., & Sörbom, D. (1988). *LISREL 7, A guide to the program and applications*. Chicago: SPSS.

Juran, J. M. (1974). Basic concepts. In J. M. Juran, F. M. Gryna, Jr., & R. S. Bingham, Jr. (Eds.), *Quality control handbook*, (3rd ed., p. 2-2). New York: McGraw-Hill.

Kelada, J. N. (1996). *Integrating reengineering with total quality*. Milwaukee, WI: American Society for Quality Control Press.

Kettinger, W. J., & Lee, C. C. (1994). Perceived service quality and user satisfaction with the information services function. *Decision Sciences*, 25(5/6), 737-766.

Kim, K. K. (1990). *User information satisfaction: Toward conceptual quality*. Paper presented at the 10th International Conference on Information Systems, Copenhagen, Denmark.

Labrecque, D. P. (1978). A response rate experiment using mail questionnaires. *Journal of Marketing*, 42(4), 82-83.

Larsen, D. L., Attkisson, C. C., Hargreaves, W. A., & Nguyen, T. D. (1979). Assessment of client/patient satisfaction: Development of a general scale. *Evaluation and Program Planning*, 2, 197-207.

Lawley, D. N., & Maxwell, A. E. (1963). *Factor analysis as a statistical method*. London: Butterworths.

Lengnick-Hall, C. A. (1996). Customer contributions to quality: A different view of the customer-oriented firm. *Academy of Management Review*, 21(3), 701-824.

- Lethinen, J. R., & Lethinen, U. (1982). *Service Quality: a Study of Quality Dimensions* (Working paper). Helsinki: Service Management Institute.
- Lineberry, R. L. (1977). Introduction: On the politics and economics of urban services. *Urban Affairs Quarterly*, 12(3), 267-272.
- Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York: John Wiley & Sons.
- Lord, F. M. (1963). Elementary models for measuring change. In C. W. Harris (Ed.), *Problems in Measuring Change*, (pp. 21-38). Madison, WI: University of Wisconsin Press.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Lovrich, N. P., Jr., & Taylor, G. T., Jr. (1976). Neighborhood evaluation of local government services: A citizen survey approach. *Urban Affairs Quarterly*, 12(December), 197-222.
- Lyons, W. E., Lowery, D., & DeHoog, R. H. (1992). *The politics of dissatisfaction: Citizens, services, and urban institutions*. Armonk, NY: M. E. Sharpe.
- Manrai, L. A. (1993). Mood effects in services: An integrated conceptual model. In T. A. Swartz, D. E. Bowen, & S. W. Brown (Eds.), *Advances in service marketing and management: Research and practice*, (Vol. 2, pp. 151-174). Greenwich, CT: JAI Press.
- Marans, R. W., & Rodgers, W. (1975). Towards an understanding of community satisfaction. In A. H. Hawley & V. P. Rock (Eds.), *Metropolitan America in contemporary perspective*, (pp. 299-352). New York: Halsted Press.
- Marsh, H. W. (1987). The hierarchical structure of self-concept and the application of hierarchical confirmatory factor analysis. *Journal of Educational Measurement*, 24(1), 17-39.
- Marsh, H. W. (1995). Δ^2 and χ^2/df fit indices for structural equation models: A brief note of clarification. *Structural Equation Modeling*, 2(3), 246-254.
- Marsh, H. W., & Balla, J. R. (1994). Goodness-of-fit in confirmatory factor analysis: The effects of sample size and model parsimony. *Quality & Quantity*, 28, 185-217.
- Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin*, 103(3), 391-410.
- Marsh, H. W., & Grayson, D. (1994). Longitudinal stability of latent means and individual differences: A unified approach. *Structural Equation Modeling*, 1(4), 317-359.
- Mason, G. (1996). Recent advances in questionnaire design for program evaluation. *Canadian Journal of Program Evaluation*, 11(1), 73-94.

McClendon, M. J., & O'Brien, D. J. (1988). Explaining question-order effects in the relationship between safety and neighborhood satisfaction. *Social Science Quarterly*, 69(3), 764-772.

Mels, G., Boshoff, C., & Nel, D. (1997). The dimensions of service quality: The original European perspective revisited. *Service Industries Journal*, 17(1), 173-189.

Moore, H. E. (1994). *Customer perception of quality in continuing education*. Unpublished doctoral dissertation, University of Alabama, Tuscaloosa.

Morrison, D. F. (1976). *Multivariate statistical methods*, (2nd ed., pp. 37-75). New York: McGraw-Hill.

Mueller, R. O. (1995). *Basic principles of structural equation modeling: An introduction to LISREL and EQS*. New York: Springer-Verlag.

Mulaik, S. A. (1972). *The foundations of factor analysis*. New York: McGraw-Hill.

Mulaik, S. A., James, L. R., van Alstine, J., Bennett, N., Lind, S., & Stilwell, C. D. (1989). Evaluation of goodness-of-fit indices for structural equation models. *Psychological Bulletin*, 105(3), 430-445.

Murdock, S. H., & Schriener, E. C. (1979). Community services satisfaction and stages of community development: An examination of evidence from impacted communities. *Journal of the Community Development Society*, 10(1), 109-124.

Nelson, P. (1970). Information and consumer behavior. *Journal of Political Economy*, 78(2), 311-339.

Norusis, M. J. (1988). *SPSS-X™ Advanced statistics guide*. (2nd ed.). Chicago: SPSS.

Nunnally, J. C. (1967). *Psychometric theory*. New York: McGraw-Hill.

Nunnally, J. C. (1978). *Psychometric theory*. (2nd ed.). New York: McGraw-Hill.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*. (3rd ed.). New York: McGraw-Hill.

Olivier, R. L. (1981). Measurement and evaluation of satisfaction processes in retail settings. *Journal of Retailing*, 57(3), 25-48.

Olivier, R. L. (1993). A conceptual model of service quality and service satisfaction: Compatible goals, different concepts. In T. A. Swartz, D. E. Bowen, & S. W. Brown (Eds.), *Advances in service marketing and management: Research and practice*, (Vol. 2, pp. 65-85). Greenwich, CT: JAI Press.

Orwig, R. A. (1994). *An empirical analysis of service quality in the Air National Guard*. Unpublished doctoral dissertation, Mississippi State University.

Parasuraman, A., Berry, L. L., & Zeithaml, V. A. (1991a). Refinement and reassessment of SERVQUAL scale. *Journal of Retailing*, 67(4), 420-450.

Parasuraman, A., Berry, L. L., & Zeithaml, V. A. (1991b). Understanding customer expectation of service. *Sloan Management Review*, 32(3), 39-48.

Parasuraman, A., Berry, L. L., & Zeithaml, V. A. (1993). More on improving service quality measurement. *Journal of Retailing*, 69(1), 140-147.

Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1985). A conceptual model of service quality and its implications for future research. *Journal of Marketing*, 49(Fall), 41-50.

Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1988). SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality. *Journal of Retailing*, 64(1), 12-40.

Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1994a). *Moving forward in service quality research: Measuring different customer-expectation levels, comparing alternative scales, and examining the performance-behavioral intentions link* (Working Paper, 94-114). Cambridge, MA: Marketing Science Institute.

Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1994b). Reassessment of expectations as a comparison standard in measuring service quality: Implications for further research. *Journal of Marketing*, 58(January), 111-124.

Parks, R. B. (1984). Linking objective and subjective measures of performance. *Public Administration Review*, 44(March/April), 118-127.

Pelissero, J. (1978). *Citizen evaluations of community services in Oklahoma*. Norman: University of Oklahoma, Bureau of Government Research.

Percy, S. L. (1986). In defence of citizen evaluations as performance measures. *Urban Affairs Quarterly*, 22(1), 66-83.

Peter, J. P., Churchill, G. A., Jr., & Brown, T. J. (1993). Caution in the use of difference scores in consumer research. *Journal of Consumer Research*, 19(March), 655-662.

Peterson, R. A., & Wilson, W. R. (1992). Measuring consumer satisfaction: Fact and artifact. *Journal of the Academy of Marketing Science*, 20(1), 61-71.

Pettigrew, D. (1993). *Qualité vs satisfaction dans les services professionnels: Le cas des experts-comptables au Québec*. Unpublished doctoral dissertation, Université de Droit, d'Économie et des Sciences d'Aix-Marseille, Aix-en-Provence, France.

Picherak, J. R. (1987). Service delivery and client satisfaction in the public sector. *Canadian Public Administration*, 30(2), 243-254.

Poister, T. H., & Henry, G. T. (1994). Citizen ratings of public and private service quality: A comparative perspective. *Public Administration Review*, 54(2), 155-160.

Pollitt, C., & Bouckaert, G. (1995). Defining quality. In C. Pollitt & G. Bouckaert (Eds.), *Quality improvements in European public services: Concepts, cases and commentary*, (pp. 20-28). London: Sage.

- Quiram, B. J. (1995). *Use of the SERVQUAL scale to measure expectations and perceptions of service quality in an ambulatory health care setting*. Unpublished doctoral dissertation, Texas A&M University, College Station.
- Rahim, M. A., & Magner, N. R. (1995). Confirmatory factor analysis of the styles of handling interpersonal conflict: First-order factor model and its invariance across groups. *Journal of Applied Psychology*, 80(1), 122-132.
- Raykov, T., & Widaman, K. F. (1995). Issues in applied structural equation modeling research. *Structural Equation Modeling*, 2(4), 289-318.
- Reeves, C. A., & Bednar, D. A. (1994). Defining quality: Alternatives and implications. *Academy of Management Review*, 19(3), 419-445.
- Richard, M. D., & Allaway, A. W. (1993). Service quality attributes and choice behavior. *Journal of Services Marketing*, 7(1), 59-68.
- Rigdon, E. E. (1995). A necessary and sufficient identification rule for structural models estimated in practice. *Multivariate Behavioral Research*, 30(3), 359-383.
- Rigdon, E. E. (1996). CFI versus RMSEA: A comparison of two fit indices for structural equation modeling. *Structural Equation Modeling*, 3(4), 369-379.
- Rindskopf, D., & Rose, T. (1988). Some theory and application of confirmatory second-order factor analysis. *Multivariate Behavioral Research*, 23(January), 51-67.
- Rosen, L. D., & Karwan, K. R. (1994). Prioritizing the dimensions of service quality: An empirical investigation and strategic assessment. *International Journal of Service Industry Management*, 5(4), 39-52.
- Rosentraub, M. S., & Thompson, L. (1981). The use of surveys of satisfaction for evaluations. *Policy Studies Journal*, 9(7), 990-999.
- Roth, V. J., Bozinoff, L., & MacIntosh, P. (1990). Public opinion and the measurement of consumer satisfaction with government services. *Canadian Public Administration*, 33(4), 571-583.
- Satorra, A., & Bentler, P. M. (1988, August 22-25). *Scaling corrections for chi-square statistics in covariance structure analysis*. Paper presented at the Annual Meeting of the American Statistical Association, New Orleans, LA.
- Scheaffer, R. L., Mendenhall, W., & Ott, R. L. (1996). *Elementary survey sampling*. (5th ed.). Belmont, CA: Duxbury Press.
- Schumacker, R. E., & Lomax, R. G. (1996). *A beginners' guide to structural equation modeling*. Mahwah, NJ: Lawrence Erlbaum.
- Schuman, H., & Gruenberg, B. (1972). Dissatisfaction with city services: Is race an important factor? In H. Hahn (Ed.), *People and politics in urban society*, (pp. 369-392). Beverly Hills, CA: Sage.
- Schwartz, N., & Hippler, H.-J. (1995). Subsequent questions may influence answers to preceding questions in mail surveys. *Public Opinion Quarterly*, 59, 93-97.

Segars, A. H. (1997). Assessing the unidimensionality of measurement: A paradigm and illustration within the context of information systems research. *Omega, International Journal of Management Science*, 25(1), 107-121.

Séguin, F. (1991). Services to the public: A major strategic change. *Canadian Public Administration*, 34(3), 465-473.

Spreng, R. A., & Singh, A. K. (1993). An empirical assessment of the SERVQUAL scale and the relationship between service quality and satisfaction. In D. W. Cravens & P. R. Dickson (Eds.), *Enhancing knowledge development in marketing: AMA educators' proceedings*, (Vol. 4, pp. 1-6). Chicago: American Marketing Association.

SPSS. (1994). *SPSS® 6.1 Syntax reference guide*. Chicago: SPSS.

Statistics Canada. (1991a). *Census handbook* (92-305E). Ottawa, Canada: Statistics Canada.

Statistics Canada. (1991b). *Profile of census tracts - Parts A & B* (95-325 & 95-326). Ottawa, Canada: Statistics Canada.

Steenkamp, J.-B. E. M., & van Tripp, H. C. M. (1991). The use of LISREL in validating marketing constructs. *International Journal of Research in Marketing*, 8(3), 283-299.

Steiger, J. H., & Lind, J. C. (1980, May 30). *Statistically-based tests for the number of common factors*. Paper presented at the Spring Meeting of the Psychometrika Society, Iowa City, IA.

Stipak, B. (1974). *Citizens' Evaluation of Municipal Services in Los Angeles County*. Los Angeles: UCLA, Institute of Government and Public Affairs.

Stipak, B. (1979a). Are there sensible ways to analyze and use subjective indicators of urban service quality? *Social Indicators Research*, 6(October), 421-438.

Stipak, B. (1979b). Citizen satisfaction with urban services: Potential misuse as a performance indicator. *Public Administration Review*, 39(January/February), 46-52.

Swartz, T. A., & Brown, S. W. (1989). Consumer and provider expectations and experiences in evaluating professional service quality. *Journal of the Academy of Marketing Science*, 17(2), 189-195.

Tabachnick, B. G., & Fidell, L. S. (1989). *Using multivariate statistics*. (2nd ed.). New York: HarperCollins.

Tan, G., & Murrell, S. (1984). Respondent characteristics in citizen evaluation of municipal services. *Social Indicators Research*, 14(January), 29-52.

Tan, T. S. W., & Barnes, D. J. (1996, May 15-17). *Measuring service quality of a regional museum: Using modified SERVQUAL model*. Paper presented at the New Zealand Organization for Quality Annual Conference, Dunedin, NZ.

Taylor, G. S. (1982). *Class and equity in public service delivery in Kentucky*. Unpublished doctoral dissertation, University of Kentucky, Lexington.

- Teas, K. R. (1993). Expectations, performance evaluation, and consumers' perceptions of quality. *Journal of Marketing*, 57(October), 18-34.
- Thompson, K. N., & Getty, J. M. (1994). Structural model of relations among quality, satisfaction, and recommending behavior in lodging decisions. *Structural Equation Modeling*, 1(2), 146-160.
- Vehorn, C. L. (1980). Substitution between public and private goods: An overview of the market meeting consumer preferences. In K. B. Monroe (Ed.), *Advances in Consumer Research*, (Vol. 8, pp. 523-526). Ann Arbor, MI: Association for Consumer Research.
- Verba, S. (1996). The citizen as respondent: Sample surveys and American democracy. *American Political Science Review*, 90(1), 1-7.
- Vogt, W. P. (1993). *Dictionary of statistics and methodology*. Newbury Park, CA: Sage.
- Waddell, H. (1995). Getting a straight answer. *Marketing Research*, 7(3), 5-8.
- Wald, A. (1950). A note on the identification of economic relations. In T. C. Koopmans (Ed.), *Statistical inference in dynamic economic models*, (pp. 238-244). New York: Wiley.
- Waldman, D. A., & Gopalakrishnan, M. (1996). Operational, organizational, and human resource factors predictive of customer perception of service quality. *Journal of Quality Management*, 1(1), 91-107.
- Wall, T. D., & Payne, R. (1973). Are deficiency scores deficient? *Journal of Applied Psychology*, 58(3), 322-326.
- Wang, L., Fan, X., & Wilson, V. L. (1996). Effects of nonnormal data on parameter estimates and fit indices for a model with latent and manifest variables: An empirical study. *Structural Equation Modeling*, 3(3), 228-247.
- Ware, W. B., Galassi, J. P., & Dew, K. M. H. (1990). The test anxiety inventory: A confirmatory factor analysis. *Anxiety Research*, 3(3), 205-212.
- Whitman-Smithe, J. D. (1995). *Evaluating quality hospital service: The relationship of expectations and social support to patients' perceptions*. Unpublished doctoral dissertation, University of Connecticut, Storrs.
- Yokoyama, F. (1992). *Causal modeling of tourist satisfaction: Application to Michigan's coastal tourism region*. Unpublished doctoral dissertation, Michigan State University.
- Zeilanga, J. R. (1994). *Privatization of auxiliary services at public universities: Advantages and disadvantages*. Unpublished doctoral dissertation, University of Arizona.
- Zeithaml, V. A. (1988). Consumer perceptions of price, quality and value: A conceptual model and synthesis of research. *Journal of Marketing*, 52(July), 2-22.

Zeithaml, V. A., Berry, L. L., & Parasuraman, A. (1993). The nature and determinants of customers expectations of service. *Journal of the Academy of Marketing Science*, 21(1), 1-12.

Zeithaml, V. A., Berry, L. L., & Parasuraman, A. (1996). The behavioral consequences of service quality. *Journal of Marketing*, 60(April), 31-46.

Zeithaml, V. A., Parasuraman, A., & Berry, L. L. (1990). *Delivering quality services: Balancing customer perceptions and expectations*. New York: The Free Press.

APPENDIX A

Determinants of Satisfaction with Municipal Services

	Schuman & Gruenberg (1972)	Stipak (1974)	Lovrich & Taylor (1976)	Pelissero (1978)
	Several	Several	Several	General, police
Socio-Demographic				
Race		-	-	M
Age		M		
Gender		M		
Income		No	No	
Education		No		
Social class status				
Occupation				
Home ownership		M		
Longitudinal				
Residence Area				
Neighbourhoods	M		+	+
Community development				
Type of dwelling				+
Size of jurisdiction	+	-		
Political Attitudes				
Social investment				
General assessment				
Political efficacy		+		
Interest in government		-		
Classes of Services				
Salience to the public				
Experience/homogeneity				
Discretion to agent				
Objective Measures				
Usage		M		-
Performance				
Subjective Measures				
Safety				+
Promptness				
Availability				
Quality				

Note. + = significant and positive findings; - = significant and negative findings; N.L. = significant but non-linear findings; M = mixed findings; No = non-significant findings

^aIncludes typical municipal services, but also non-municipal services (schools, mental health), as well as commercial services (shopping, housing availability)

	Murdock & Schriner (1979) Com.services ^a	Goodsell (1981) Welfare	Rosentraub & Thompson (1981) Several	Brown & Coulter (1983) Police
Socio-Demographic				
Race			No	-
Age	+	+		+
Gender				+
Income			No	+
Education				No
Social class status				
Occupation				
Home ownership				
Longitudinal				
Residence Area				
Neighbourhoods				
Community development	N.L.			
Type of dwelling				
Size of jurisdiction				
Political Attitudes				
Social investment				
General assessment				+
Political efficacy				+
Interest in government				
Classes of Services				
Salience to the public				
Experience/homogeneity				
Discretion to agent			-	
Objective Measures				
Usage				M
Performance				No
Subjective Measures				
Safety				+
Promptness				
Availability				
Quality				

Note. + = significant and positive findings; - = significant and negative findings; N.L. = significant but non-linear findings; M = mixed findings; No = non-significant findings

^aIncludes typical municipal services, but also non-municipal services (schools, mental health), as well as commercial services (shopping, housing availability)

	Parks (1984)	Tan & Murrell (1984)	Brede (1985)	Glauser & Tullar (1985)
	Police	Several	Military police	Police
Socio-Demographic				
Race		M	No	
Age				No
Gender				No
Income				
Education				No
Social class status			+	
Occupation				
Home ownership				
Longitudinal		No		
Residence Area				
Neighbourhoods			No	
Community development				
Type of dwelling				
Size of jurisdiction				
Political Attitudes				
Social investment				
General assessment				
Political efficacy				
Interest in government				
Classes of Services				
Salience to the public				
Experience/homogeneity				
Discretion to agent				
Objective Measures				
Usage				
Performance		No	No	
Subjective Measures				
Safety				
Promptness	+			
Availability				
Quality				

Note. + = significant and positive findings; - = significant and negative findings; N.L. = significant but non-linear findings; M = mixed findings; No = non-significant findings

^aIncludes typical municipal services, but also non-municipal services (schools, mental health), as well as commercial services (shopping, housing availability)

	Percy (1986)	McClendon & O'Brien (1988)	Roth, Bozinoff, & MacIntosh (1990)	Lyons, Lowery, & DeHoog (1992)
	Police	Generalized	Several	Generalized
Socio-Demographic				
Race				
Age	+			
Gender			No	No
Income			M	
Education			No	
Social class status				
Occupation			No	
Home ownership				No
Longitudinal				
Residence Area				
Neighbourhoods				
Community development				
Type of dwelling				
Size of jurisdiction			+	
Political Attitudes				
Social investment				+
General assessment				
Political efficacy				+
Interest in government				
Classes of Services				
Salience to the public			-	
Experience/homogeneity				
Discretion to agent				
Objective Measures				
Usage				+
Performance				
Subjective Measures				
Safety	+	M		
Promptness	+			
Availability				
Quality				+

Note. + = significant and positive findings; - = significant and negative findings; N.L. = significant but non-linear findings; M = mixed findings; No = non-significant findings

^aIncludes typical municipal services, but also non-municipal services (schools, mental health), as well as commercial services (shopping, housing availability)

Das, Das, & McKenzie (1995)		
Several		
Socio-Demographic		
Race		
Age	No	
Gender		
Income	No	
Education	No	
Social class status		
Occupation		
Home ownership		
Longitudinal		
Residence Area		
Neighbourhoods		
Community development		
Type of dwelling		
Size of jurisdiction		
Political Attitudes		
Social investment		
General assessment		
Political efficacy		
Interest in government		
Classes of Services		
Salience to the public		
Experience/homogeneity	-	
Discretion to agent		
Objective Measures		
Usage		
Performance		
Subjective Measures		
Safety		
Promptness		
Availability		
Quality		

Note. + = significant and positive findings; - = significant and negative findings; N.L. = significant but non-linear findings; M = mixed findings; No = non-significant findings

^aIncludes typical municipal services, but also non-municipal services (schools, mental health), as well as commercial services (shopping, housing availability)

APPENDIX B**Survey Material**

Appendix B1: English questionnaire

Appendix B2: French questionnaire

Appendix B3: Transmittal letter



Section 1 : À partir de votre expérience de résidant, pensez à une municipalité qui fournirait des services d'excellente qualité. Pensez à une municipalité où vous seriez heureux d'habiter à cause des services de très grande qualité que vous y recevriez. Veuillez indiquer à quel degré cette municipalité posséderait chacune des caractéristiques énumérées dans les énoncés ci-dessous. Si vous croyez que la caractéristique n'est *pas du tout* essentielle pour une excellente municipalité, encerclez le «1». Si vous êtes d'opinion que la caractéristique est *tout à fait* essentielle pour cette excellente municipalité, encerclez le «7». Si vous avez une autre opinion, encerclez un des numéros entre le «1» et le «7». Il n'y a pas de bonnes ou de mauvaises réponses. Nous désirons seulement que vous indiquiez les numéros qui correspondent le mieux à vos opinions concernant une municipalité qui fournirait des services d'excellente qualité.

		<u>Pas du tout d'accord</u>				<u>Tout à fait d'accord</u>			
		1	2	3	4	5	6	7	
1	Une excellente municipalité possède un équipement récent.	1	2	3	4	5	6	7	
2	Les installations matérielles d'une excellente municipalité sont visuellement attrayantes.	1	2	3	4	5	6	7	
3	Les employés d'une excellente municipalité ont une apparence soignée.	1	2	3	4	5	6	7	
4	Les supports visuels associés au service (tels que les dépliants ou les affiches) dans une excellente municipalité sont attirants.	1	2	3	4	5	6	7	
5	Quand une excellente municipalité s'engage à réaliser quelque chose, elle s'y tient.	1	2	3	4	5	6	7	
6	Quand les résidants rencontrent un problème, une excellente municipalité essaie de le résoudre.	1	2	3	4	5	6	7	
7	Une excellente municipalité rend le bon service du premier coup.	1	2	3	4	5	6	7	
8	Une excellente municipalité fournit ses services dans les délais convenus.	1	2	3	4	5	6	7	
9	Une excellente municipalité gère ses fichiers et registres sans erreurs.	1	2	3	4	5	6	7	
10	Les employés d'une excellente municipalité disent aux résidants exactement quand le service sera fourni.	1	2	3	4	5	6	7	
11	Les employés d'une excellente municipalité assurent un service prompt.	1	2	3	4	5	6	7	
12	Les employés d'une excellente municipalité sont toujours disposés à aider.	1	2	3	4	5	6	7	
13	Les employés d'une excellente municipalité ne sont jamais trop occupés pour répondre aux demandes des résidants.	1	2	3	4	5	6	7	
14	On peut faire confiance aux employés d'une excellente municipalité.	1	2	3	4	5	6	7	
15	Il est rassurant de faire affaires avec une excellente municipalité.	1	2	3	4	5	6	7	
16	Les employés d'une excellente municipalité sont toujours polis.	1	2	3	4	5	6	7	
17	Les employés d'une excellente municipalité ont les connaissances nécessaires pour répondre aux questions des résidants.	1	2	3	4	5	6	7	
18	Une excellente municipalité donne aux résidants une attention individuelle.	1	2	3	4	5	6	7	

	<u>Pas du tout d'accord</u>				<u>Tout à fait d'accord</u>			
19 Les heures d'ouverture des services d'une excellente municipalité sont pratiques pour tous les résidants.	1	2	3	4	5	6	7	
20 Les employés d'une excellente municipalité accordent aux résidants une attention personnelle.	1	2	3	4	5	6	7	
21 Une excellente municipalité a à coeur d'agir selon les meilleurs intérêts des résidants.	1	2	3	4	5	6	7	
22 Les employés d'une excellente municipalité comprennent les besoins spécifiques des résidants.	1	2	3	4	5	6	7	

Section 2 : La liste ci-dessous énumère des caractéristiques relatives aux municipalités et aux services qu'elles fournissent. Veuillez indiquer le degré d'importance que *vous* attachez à chacune de ces caractéristiques lorsque vous évaluez la qualité des services municipaux. Répartissez le total de 100 points entre les caractéristiques *selon l'importance que vous accordez à chacune* : plus une caractéristique est importante, plus vous devriez lui accorder de points. Veuillez vous assurer que le total est bien de 100.

- 23 A - L'apparence des installations matérielles, équipements et employés de la municipalité. _____ points
- 24 B - L'habileté de la municipalité à fournir les services avec fiabilité et exactitude. _____ points
- 25 C - L'engagement de la municipalité à fournir un service prompt et à aider les résidants. _____ points
- 26 D - Des employés municipaux qui sont compétents, courtois et qui inspirent confiance. _____ points
- 27 E - L'écoute et l'attention individuelle que la municipalité accorde à ses résidants. _____ points
- Total 100 points**
- 28 Parmi les 5 caractéristiques ci-dessus, *laquelle* est la *plus importante* pour vous?
(veuillez inscrire la lettre correspondante : A, B, C, D, ou E) _____
- 29 Quelle caractéristique est la *seconde en importance*? _____
- 30 À quelle caractéristique accordez-vous le *moins d'importance*? _____

Section 3 : Les énoncés qui suivent sollicitent votre opinion à l'égard des services municipaux qui vous sont rendus par la ville de **Saint-Jean-sur-Richelieu** (dorénavant, nous utiliserons «Saint-Jean»). Comme dans la section 1, encerclez un des numéros de «1» à «7» pour indiquer précisément votre opinion.

	<u>Pas du tout d'accord</u>				<u>Tout à fait d'accord</u>			
31 Saint-Jean possède un équipement récent.	1	2	3	4	5	6	7	
32 Les installations matérielles de Saint-Jean sont visuellement attrayantes.	1	2	3	4	5	6	7	
33 Les employés de Saint-Jean ont une apparence soignée.	1	2	3	4	5	6	7	
34 Les supports visuels associés au service (tels que les dépliants ou les affiches) à Saint-Jean sont attirants.	1	2	3	4	5	6	7	
35 Quand Saint-Jean s'engage à réaliser quelque chose, elle s'y tient.	1	2	3	4	5	6	7	
36 Quand vous rencontrez un problème, Saint-Jean essaie de le résoudre.	1	2	3	4	5	6	7	

Passez à la page suivante →

	<u>Pas du tout d'accord</u>				<u>Tout à fait d'accord</u>			
37 Saint-Jean rend le bon service du premier coup.	1	2	3	4	5	6	7	
38 Saint-Jean fournit ses services dans les délais convenus.	1	2	3	4	5	6	7	
39 Saint-Jean gère ses fichiers et registres sans erreurs.	1	2	3	4	5	6	7	
40 Les employés de Saint-Jean vous disent exactement quand le service sera fourni.	1	2	3	4	5	6	7	
41 Les employés de Saint-Jean vous assurent un service prompt.	1	2	3	4	5	6	7	
42 Les employés de Saint-Jean sont toujours disposés à vous aider.	1	2	3	4	5	6	7	
43 Les employés de Saint-Jean ne sont jamais trop occupés pour vous servir.	1	2	3	4	5	6	7	
44 Vous pouvez faire confiance aux employés de Saint-Jean.	1	2	3	4	5	6	7	
45 Vous êtes rassuré de faire affaires avec Saint-Jean.	1	2	3	4	5	6	7	
46 Les employés de Saint-Jean sont toujours polis avec vous.	1	2	3	4	5	6	7	
47 Les employés de Saint-Jean ont les connaissances nécessaires pour répondre à vos questions.	1	2	3	4	5	6	7	
48 Saint-Jean vous accorde une attention individuelle.	1	2	3	4	5	6	7	
49 Les heures d'ouverture des services de Saint-Jean sont pratiques pour vous.	1	2	3	4	5	6	7	
50 Les employés de Saint-Jean vous accordent une attention personnelle.	1	2	3	4	5	6	7	
51 Saint-Jean a à coeur d'agir selon vos meilleurs intérêts.	1	2	3	4	5	6	7	
52 Les employés de Saint-Jean comprennent vos besoins spécifiques.	1	2	3	4	5	6	7	

Section 4 : À titre de résidant et de consommateur de services municipaux, donnez votre opinion concernant votre satisfaction à l'égard des services municipaux de Saint-Jean. Comme dans la section 1, encerclez un des numéros de «1» à «7» pour indiquer précisément votre opinion.

	<u>Pas du tout d'accord</u>				<u>Tout à fait d'accord</u>			
53 Les résidents des autres municipalités de la région du Haut-Richelieu sont plus satisfaits des services municipaux qu'ils reçoivent que je ne le suis.	1	2	3	4	5	6	7	
54 Je suis entièrement satisfait des services municipaux de Saint-Jean.	1	2	3	4	5	6	7	
55 À cause de la mauvaise qualité des services municipaux de Saint-Jean, j'envisage de déménager dans une autre ville.	1	2	3	4	5	6	7	
56 Presque tous les résidents sont satisfaits des services municipaux de Saint-Jean.	1	2	3	4	5	6	7	
57 Je recommanderais Saint-Jean comme une ville où les résidents sont très satisfaits des services municipaux reçus.	1	2	3	4	5	6	7	

Passez à la page suivante ➡

Cette section n'est pas reliée à l'étude. Elle est incluse seulement à titre de source d'information pour la Chambre de commerce du Haut-Richelieu.

Section 5 : En vous appuyant sur votre connaissance du milieu régional, veuillez donner votre opinion sur les énoncés suivants touchant la Chambre de commerce du Haut-Richelieu (dorénavant, nous utiliserons «Chambre de commerce»). Comme dans la section 1, encerclez un des numéros de «1» à «7» pour indiquer précisément votre opinion.

	<u>Pas du tout d'accord</u>				<u>Tout à fait d'accord</u>			
58 Je connais bien le rôle de la Chambre de commerce.	1	2	3	4	5	6	7	
59 La Chambre de commerce est le porte-parole du monde des affaires de la région du Haut-Richelieu.	1	2	3	4	5	6	7	
60 La Chambre de commerce fournit des services aux entreprises.	1	2	3	4	5	6	7	
61 La Chambre de commerce facilite la mise en oeuvre de projets à caractère économique.	1	2	3	4	5	6	7	
62 La Chambre de commerce est le moteur de l'économie régionale.	1	2	3	4	5	6	7	

Section 6 : Veuillez cocher la case correspondant à votre réponse pour chacune des questions suivantes.

- | | |
|--|--|
| <p>63 Veuillez indiquer votre statut d'emploi.</p> <p>1. Travail à temps plein <input type="checkbox"/></p> <p>2. Travail à temps partiel <input type="checkbox"/></p> <p>3. En recherche d'emploi <input type="checkbox"/></p> <p>4. Retraité <input type="checkbox"/></p> <p>5. Autre <input type="checkbox"/></p> | <p>67 Est-ce que vous ou un autre membre du ménage êtes le propriétaire de ce logement (même s'il y a une hypothèque) ou est-ce que vous êtes locataire (même sans payer de loyer)?</p> <p>1. Propriétaire <input type="checkbox"/></p> <p>2. Locataire <input type="checkbox"/></p> |
| <p>64 Sexe.</p> <p>m. Masculin <input type="checkbox"/></p> <p>f. Féminin <input type="checkbox"/></p> | <p>68 Veuillez indiquer votre revenu familial total.</p> <p>1. Moins de 10 000 \$ <input type="checkbox"/></p> <p>2. 10 000 \$ à 19 999 \$ <input type="checkbox"/></p> <p>3. 20 000 \$ à 29 999 \$ <input type="checkbox"/></p> <p>4. 30 000 \$ à 39 999 \$ <input type="checkbox"/></p> <p>5. 40 000 \$ à 49 999 \$ <input type="checkbox"/></p> <p>6. 50 000 \$ à 59 999 \$ <input type="checkbox"/></p> <p>7. 60 000 \$ ou plus <input type="checkbox"/></p> |
| <p>65 Veuillez indiquer le plus haut niveau de scolarité que vous ayez complété.</p> <p>1. École primaire <input type="checkbox"/></p> <p>2. École secondaire <input type="checkbox"/></p> <p>3. CEGEP ou équivalent <input type="checkbox"/></p> <p>4. Université <input type="checkbox"/></p> | <p>69 Veuillez indiquer votre groupe d'âge.</p> <p>1. 0 à 24 ans <input type="checkbox"/></p> <p>2. 25 à 34 ans <input type="checkbox"/></p> <p>3. 35 à 44 ans <input type="checkbox"/></p> <p>4. 45 à 54 ans <input type="checkbox"/></p> <p>5. 55 à 64 ans <input type="checkbox"/></p> <p>6. 65 ans ou plus <input type="checkbox"/></p> |
| <p>66 Depuis combien de temps demeurez-vous à Saint-Jean?</p> <p>1. Moins d'un an <input type="checkbox"/></p> <p>2. 1 à 4 ans <input type="checkbox"/></p> <p>3. 5 ans ou plus <input type="checkbox"/></p> | |

Encore une fois merci de votre participation au sondage. Si vous avez des commentaires à l'égard de la qualité des services municipaux, veuillez les joindre au questionnaire.



The Quality of Municipal Services and Residents' Satisfaction

Section 1: Based on your experiences as a resident, please think about what kind of municipality would deliver excellent quality of services. Think about the kind of municipality from which you would be pleased to receive services. Please indicate the extent to which you think such a municipality would possess the feature described by each of the following statements. If you feel a feature is *not at all essential* for an excellent municipality, circle the number "1". If you feel a feature is *absolutely essential* for an excellent municipality, circle "7". If your feelings are not as strong, circle one of the numbers in between. There are no right or wrong answers--all we are interested in is a number that truly reflects your feelings regarding municipalities that would deliver excellent quality of service.

		Strongly disagree					Strongly agree	
		1	2	3	4	5	6	7
1	Excellent municipalities will have modern equipment.	1	2	3	4	5	6	7
2	Physical facilities at excellent municipalities will be visually appealing.	1	2	3	4	5	6	7
3	Employees of excellent municipalities will have a neat appearance.	1	2	3	4	5	6	7
4	Visual materials associated with service to residents (such as pamphlets or correspondence) will be appealing in excellent municipalities.	1	2	3	4	5	6	7
5	When excellent municipalities promise to do something, they will do it.	1	2	3	4	5	6	7
6	When residents have a problem, excellent municipalities will show a sincere interest in solving it.	1	2	3	4	5	6	7
7	Excellent municipalities will perform the service right the first time.	1	2	3	4	5	6	7
8	Excellent municipalities will provide their services at the time they promise to do so.	1	2	3	4	5	6	7
9	Excellent municipalities will insist on error-free records.	1	2	3	4	5	6	7
10	Employees of excellent municipalities will tell residents exactly when services will be performed.	1	2	3	4	5	6	7
11	Employees of excellent municipalities will give prompt service.	1	2	3	4	5	6	7
12	Employees of excellent municipalities will always be willing to help.	1	2	3	4	5	6	7
13	Employees of excellent municipalities will never be too busy to respond to residents' requests.	1	2	3	4	5	6	7
14	The behavior of employees of excellent municipalities will instill confidence in residents.	1	2	3	4	5	6	7
15	It is comforting to have dealings with excellent municipalities.	1	2	3	4	5	6	7
16	Employees of excellent municipalities will be consistently courteous towards residents.	1	2	3	4	5	6	7
17	Employees of excellent municipalities will have the knowledge to answer residents' questions.	1	2	3	4	5	6	7
18	Excellent municipalities will give residents individual attention.	1	2	3	4	5	6	7

Turn to the next page ⇒

	<u>Strongly disagree</u>					<u>Strongly agree</u>	
19	1	2	3	4	5	6	7
Excellent municipalities will have operating hours that are convenient for all their residents.							
20	1	2	3	4	5	6	7
Excellent municipalities will have employees who give residents personal attention.							
21	1	2	3	4	5	6	7
Excellent municipalities will have the residents' best interest at heart.							
22	1	2	3	4	5	6	7
The employees of excellent municipalities will understand the specific needs of their residents.							

Section 2: Below are five features pertaining to municipalities and their services. We would like to know how important each feature is to you when evaluating the quality of municipal services. Please allocate a total of 100 points among the five features according to their importance to you: the more important a feature is to you, the more points you should allocate to it. Please ensure that the points allocated to the five features add up to 100.

- 23 A - The appearance of the municipality's physical facilities, equipment and personnel. _____ points
- 24 B - The municipality's ability to perform the promised service dependably and accurately. _____ points
- 25 C - The municipality's willingness to help residents and provide prompt service. _____ points
- 26 D - The knowledge and courtesy of municipal employees and their ability to convey trust. _____ points
- 27 E - The caring and individualized attention the municipality provides for its residents. _____ points
- Total : 100 points**

28 Which *one* of the above 5 features is *most important* to you? (please enter the corresponding letter: A, B, C, D, or E) _____

29 Which feature is *second most important* to you? _____

30 Which feature is *least important* to you? _____

Section 3: The following set of statements relates to your feelings about the municipal services of the town of Saint-Jean-sur-Richelieu (referred to hereafter as "Saint-Jean"). As in Section 1, you may circle any of the numbers from "1" to "7" to indicate how strong your feelings are.

	<u>Strongly disagree</u>					<u>Strongly agree</u>	
31	1	2	3	4	5	6	7
Saint-Jean has modern equipment.							
32	1	2	3	4	5	6	7
Saint-Jean's physical facilities are visually appealing.							
33	1	2	3	4	5	6	7
Saint-Jean's employees have a neat appearance.							
34	1	2	3	4	5	6	7
Visual materials associated with service to residents (such as pamphlets or correspondence) are appealing in Saint-Jean.							
35	1	2	3	4	5	6	7
When Saint-Jean promises to do something, it does it.							
36	1	2	3	4	5	6	7
When you have a problem, Saint-Jean shows a sincere interest in solving it.							

Turn to the next page ⇒

	<u>Strongly disagree</u>				<u>Strongly agree</u>			
37 Saint-Jean performs the service right the first time.	1	2	3	4	5	6	7	
38 Saint-Jean provides its services at the time it promises to do so.	1	2	3	4	5	6	7	
39 Saint-Jean insists on error-free records.	1	2	3	4	5	6	7	
40 Employees of Saint-Jean tell you exactly when services will be performed.	1	2	3	4	5	6	7	
41 Employees of Saint-Jean give you prompt service.	1	2	3	4	5	6	7	
42 Employees of Saint-Jean are always willing to help you.	1	2	3	4	5	6	7	
43 Employees of Saint-Jean are never too busy to respond to your requests.	1	2	3	4	5	6	7	
44 The behavior of the employees of Saint-Jean instills confidence in you.	1	2	3	4	5	6	7	
45 It is comforting for you to have dealings with Saint-Jean.	1	2	3	4	5	6	7	
46 Employees of Saint-Jean are consistently courteous towards you.	1	2	3	4	5	6	7	
47 Employees of Saint-Jean have the knowledge to answer your questions.	1	2	3	4	5	6	7	
48 Saint-Jean gives you individual attention.	1	2	3	4	5	6	7	
49 Saint-Jean has operating hours that are convenient for you.	1	2	3	4	5	6	7	
50 Saint-Jean has employees who give you personal attention.	1	2	3	4	5	6	7	
51 Saint-Jean has your best interest at heart.	1	2	3	4	5	6	7	
52 Employees of Saint-Jean understand your specific needs.	1	2	3	4	5	6	7	

Section 4: Based on your experiences as a resident and consumer of municipal services, please indicate how you feel about the following statements regarding your satisfaction with the services provided by Saint-Jean. As in Section 1, you may circle any of the numbers from "1" to "7" to indicate how strong your feelings are.

	<u>Strongly disagree</u>				<u>Strongly agree</u>			
53 Persons living in other municipalities of the Haut-Richelieu region are more satisfied with the municipal services they receive than I am.	1	2	3	4	5	6	7	
54 I am fully satisfied with the services provided by Saint-Jean.	1	2	3	4	5	6	7	
55 Because of the poor quality of municipal services in Saint-Jean, I am thinking of moving to another town.	1	2	3	4	5	6	7	
56 Almost all residents are satisfied with the municipal services of Saint-Jean.	1	2	3	4	5	6	7	
57 I would recommend St-Jean as a town where residents are highly satisfied with the municipal services provided.	1	2	3	4	5	6	7	

Turn to the next page ⇒

The following section is not related to the study. It is enclosed only to provide feedback for the Chamber of Commerce of Haut-Richelieu.

Section 5: Based on your knowledge of the socioeconomic environment in the Haut-Richelieu region, please indicate your opinion about the following statements concerning the Chamber of Commerce of Haut-Richelieu (referred to hereafter as "Chamber of Commerce"). As in Section 1, you may circle any of the numbers from "1" to "7" to indicate how strong your feelings are.

		Strongly disagree					Strongly agree	
		1	2	3	4	5	6	7
58	I am well aware of the role of the Chamber of Commerce.							
59	The Chamber of Commerce is the speaker for the business community in the Haut-Richelieu region.							
60	The Chamber of Commerce provides services to businesses.							
61	The Chamber of Commerce fosters the launch of economic projects.							
62	The Chamber of Commerce is the driving force for the regional economy.							

Section 6: Please check the box corresponding to your answer to each of the following questions.

<p>63 What is your current employment status?</p> <p>1. Employed full-time <input type="checkbox"/></p> <p>2. Employed part-time <input type="checkbox"/></p> <p>3. Seeking employment <input type="checkbox"/></p> <p>4. Retired <input type="checkbox"/></p> <p>5. Other <input type="checkbox"/></p>	<p>67 Do you or a member of this household own this dwelling (even if it is still being paid for) or do you rent it (even if no cash rent is paid)?</p> <p>1. Own <input type="checkbox"/></p> <p>2. Rent <input type="checkbox"/></p>
<p>64 Gender.</p> <p>m. Male <input type="checkbox"/></p> <p>f. Female <input type="checkbox"/></p>	<p>68 Please indicate your total family income.</p> <p>1. Less than \$ 10,000 <input type="checkbox"/></p> <p>2. \$ 10,000 to \$ 19,999 <input type="checkbox"/></p> <p>3. \$ 20,000 to \$ 29,999 <input type="checkbox"/></p> <p>4. \$ 30,000 to \$ 39,999 <input type="checkbox"/></p> <p>5. \$ 40,000 to \$ 49,999 <input type="checkbox"/></p> <p>6. \$ 50,000 to \$ 59,999 <input type="checkbox"/></p> <p>7. \$ 60,000 or more <input type="checkbox"/></p>
<p>65 Please indicate the highest education level that you have completed.</p> <p>1. Elementary school <input type="checkbox"/></p> <p>2. High school <input type="checkbox"/></p> <p>3. CEGEP or equivalent <input type="checkbox"/></p> <p>4. University <input type="checkbox"/></p>	<p>69 Please indicate your age group.</p> <p>1. 0 to 24 years <input type="checkbox"/></p> <p>2. 25 to 34 years <input type="checkbox"/></p> <p>3. 35 to 44 years <input type="checkbox"/></p> <p>4. 45 to 54 years <input type="checkbox"/></p> <p>5. 55 to 64 years <input type="checkbox"/></p> <p>6. 65 years or more <input type="checkbox"/></p>
<p>66 How long have you been living in Saint-Jean?</p> <p>1. Less than 1 year <input type="checkbox"/></p> <p>2. 1 to 4 years <input type="checkbox"/></p> <p>3. 5 years or more <input type="checkbox"/></p>	

Thank you again for your participation in this study. If you have any comments regarding the quality of municipal services, please attach them to this questionnaire.



Montréal, le 10 avril 1996

Résidentes, Résidents de Saint-Jean-sur-Richelieu,
Residents of Saint-Jean-sur-Richelieu,

Vous trouverez ci-joint un questionnaire sollicitant votre opinion à l'égard des services municipaux. Tout adulte (18 ans ou plus) habitant à cette adresse peut le remplir. L'objectif de cette recherche est d'étudier la qualité des services municipaux de la ville de Saint-Jean-sur-Richelieu, ainsi que la satisfaction des résidents à leur égard.

Votre résidence fait partie d'un code postal de Saint-Jean-sur-Richelieu qui a été choisi au hasard pour participer au sondage. Il ne vous faudra que quinze minutes pour répondre au questionnaire. On ne peut pas vous identifier par vos réponses. Vos réponses seront regroupées avec celles des autres participants et seulement les sommaires seront rendus publics.

Les résultats du sondage seront dévoilés vers la fin du mois de juin. Nous croyons fermement que les conclusions de la recherche seront utiles aux élus municipaux et pourront entraîner l'amélioration de la qualité des services publics.

Votre participation au sondage est importante pour le succès de l'étude. Veuillez répondre à toutes les questions et retourner le questionnaire le plus tôt possible dans l'enveloppe pré-affranchie jointe.

Merci de votre collaboration.

Claude Roy
Chercheur, Researcher

Please find enclosed a survey questionnaire soliciting your opinion regarding municipal services. Any adult (18 years or more) living at this address may complete it. The purpose of this study is to investigate the quality of the municipal services in the Town of Saint-Jean-sur-Richelieu, as well as the satisfaction of residents with them.

Your dwelling is part of a postal code in Saint-Jean-sur-Richelieu that has been selected at random to participate in the survey. It will take you only fifteen minutes to complete the questionnaire. You cannot be identified by your answers. Your answers will be grouped with those of the other participants and only summary data will be published.

The results of the survey will be made public around the end of the month of June. We strongly believe that the conclusions of the research will be useful to elected officials and may lead to improvements in the quality of the public services.

If you would prefer an English questionnaire, you may exchange this copy at the Town Hall of Saint-Jean-sur-Richelieu, 188, Jacques-Cartier St. N., during business hours.

Your participation in this survey is important to this study. Please answer all questions and return the questionnaire as soon as possible, using the postage-paid return envelope.

Thank you for your cooperation.

Jamshid Etezadi-Amoli, Ph.D.,
Professeur associé, Associate Professor

APPENDIX C

Socio-Demographic Attributes of the Respondents

Table C1

Distribution of the Employment Status Item (EMPL)

	<i>f</i>	%	Cumulative %
Employed full-time	235	61	61
Employed part-time	29	8	69
Seeking employment	21	5	74
Retired	76	20	94
Other	25	6	100

Note. Valid cases only. $N = 386$.

Table C2

Distribution of the Gender Item (SEX)

	<i>f</i>	%	Cumulative %
Male	236	61	61
Female	151	39	100

Note. Valid cases only. $N = 387$.

Table C3

Distribution of the Education Item (EDUC)

	<i>f</i>	%	Cumulative %
Elementary school	22	6	6
High school	138	36	42
CEGEP	124	32	74
University	102	26	100

Note. Valid cases only. $N = 386$.

Table C4

Distribution of the Stability of Residence Item (STAB)

	<i>f</i>	%	Cumulative %
Less than 1 year	8	2	2
1 to 4 years	42	11	13
5 years or more	339	87	100

Note. Valid cases only. *N* = 389.

Table C5

Distribution of the Homeownership Item (OWNE)

	<i>f</i>	%	Cumulative %
Own	304	78	78
Rent	86	22	100

Note. Valid cases only. *N* = 390.

Table C6

Distribution of the Family Income Item (INCO)

	<i>f</i>	%	Cumulative %
Less than \$10 000	13	4	4
\$10 000 to \$19 999	49	13	17
\$20 000 to \$29 999	64	17	34
\$30 000 to \$39 999	57	15	49
\$40 000 to \$49 999	58	15	64
\$50 000 to \$59 999	48	13	77
\$60 000 or more	86	23	100

Note. Valid cases only. *N* = 375.

Table C7

Distribution of the Age Item (AGEG)

	<i>f</i>	%	Cumulative %
0 to 24 years	13	3	3
25 to 34 years	57	15	18
35 to 44 years	96	25	43
45 to 54 years	109	28	71
55 to 64 years	71	18	89
65 years or more	42	11	100

Note. Valid cases only. $N = 388$.

APPENDIX D

Summary Descriptive Measures of the Survey Items and Difference Scores

Table D1

Descriptive Statistics for the Expectation Battery Items

	<i>n</i>	<i>M</i>	<i>SD</i>	Skewness	Kurtosis
EXP1	388	4.43	1.63	-0.45***	-0.22
EXP2	389	4.62	1.80	-0.49***	-0.65**
EXP3	385	5.56	1.52	-1.14***	0.96***
EXP4	389	5.41	1.54	-0.99***	0.63**
EXP5	387	6.24	1.31	-2.19***	4.90***
EXP6	390	6.49	1.03	-2.89***	10.21***
EXP7	388	5.93	1.29	-1.57***	3.03***
EXP8	389	6.35	1.03	-2.27***	6.60***
EXP9	388	6.26	1.11	-2.06***	5.37***
EXP10	390	6.32	1.09	-2.23***	6.25***
EXP11	390	6.30	1.06	-2.29***	7.02***
EXP12	388	6.26	1.18	-2.19***	5.67***
EXP13	389	6.03	1.34	-1.81***	3.50***
EXP14	387	6.07	1.41	-1.99***	3.97***
EXP15	390	6.40	1.07	-2.53***	8.00***
EXP16	390	6.43	1.03	-2.49***	7.73***
EXP17	388	6.25	1.16	-2.09***	5.40***
EXP18	390	5.75	1.40	-1.39***	2.15***
EXP19	390	5.99	1.29	-1.63***	3.10***
EXP20	388	5.84	1.27	-1.20***	1.71***
EXP21	387	6.49	1.00	-2.90***	10.87***
EXP22	388	6.10	1.23	-1.95***	4.57***

Note. Seven-point Likert scales (1 = *strongly disagree*, 7 = *strongly agree*). Valid cases only.

** $p < .01$, two-tailed; *** $p < .001$, two-tailed.

Table D2
Descriptive Statistics for the Perception Battery Items

	<i>n</i>	<i>M</i>	<i>SD</i>	Skewness	Kurtosis
PER1	383	4.77	1.17	-0.23	0.78 ^{††}
PER2	387	4.80	1.30	-0.36 ^{††}	-0.01
PER3	388	4.80	1.38	-0.50 ^{†††}	0.15
PER4	390	5.12	1.34	-0.74 ^{†††}	0.50 [†]
PER5	384	4.52	1.51	-0.45 ^{†††}	-0.18
PER6	388	4.49	1.62	-0.52 ^{†††}	-0.26
PER7	386	4.30	1.54	-0.45 ^{†††}	-0.23
PER8	385	4.50	1.50	-0.34 ^{††}	-0.33
PER9	374	4.69	1.42	-0.35 ^{††}	0.04
PER10	384	4.63	1.64	-0.41 ^{††}	-0.50 [†]
PER11	386	4.74	1.51	-0.46 ^{†††}	-0.22
PER12	384	4.80	1.61	-0.61 ^{†††}	-0.28
PER13	384	4.60	1.60	-0.54 ^{†††}	-0.22
PER14	385	4.85	1.58	-0.69 ^{†††}	-0.02
PER15	384	4.98	1.47	-0.66 ^{†††}	0.18
PER16	387	5.29	1.51	-0.97 ^{†††}	0.64 ^{††}
PER17	384	4.99	1.46	-0.79 ^{†††}	0.44
PER18	383	4.39	1.60	-0.36 ^{††}	-0.33
PER19	387	4.86	1.73	-0.69 ^{†††}	-0.36
PER20	386	4.45	1.61	-0.35	-0.41
PER21	385	4.62	1.54	-0.48 ^{†††}	-0.18
PER22	384	4.50	1.54	-0.48 ^{†††}	-0.18

Note. Seven-point Likert scales (1 = *strongly disagree*, 7 = *strongly agree*). Valid cases only.

[†]*p* < .05, two-tailed; ^{††}*p* < .01, two-tailed; ^{†††}*p* < .001, two-tailed.

Table D3

Descriptive Statistics for the Difference Score Variables

	<i>n</i>	<i>M</i>	<i>SD</i>	Skewness	Kurtosis
DIF1	381	0.34	1.70	0.61 ^{†††}	1.49 ^{†††}
DIF2	386	0.19	2.01	0.20	0.76 ^{†††}
DIF3	384	-0.77	1.90	-0.05	0.97 ^{†††}
DIF4	389	-0.28	1.69	-0.07	1.55 ^{†††}
DIF5	382	-1.74	1.96	0.32 [‡]	1.42 ^{†††}
DIF6	388	-2.01	1.82	-0.26 [‡]	0.26
DIF7	385	-1.64	1.78	-0.46 ^{†††}	-0.04
DIF8	384	-1.85	1.73	-0.23	0.24
DIF9	372	-1.55	1.63	-0.21	0.31
DIF10	384	-1.67	1.86	-0.26 [‡]	0.26
DIF11	386	-1.57	1.71	-0.30 [‡]	0.42
DIF12	382	-1.46	1.84	-0.22	0.72 ^{†‡}
DIF13	384	-1.42	1.97	-0.08	0.55 [‡]
DIF14	382	-1.24	1.93	-0.07	1.11 ^{†††}
DIF15	384	-1.43	1.69	-0.35 ^{†‡}	0.98 ^{†††}
DIF16	387	-1.15	1.70	-0.56 ^{†††}	1.60 ^{†††}
DIF17	382	-1.27	1.70	-0.43 ^{†††}	0.65 ^{†‡}
DIF18	383	-1.37	1.87	-0.04	0.91 ^{†††}
DIF19	387	-1.14	2.02	-0.29 [‡]	0.97 ^{†††}
DIF20	384	-1.39	1.78	-0.24	0.30
DIF21	382	-1.89	1.71	-0.38 ^{†‡}	-0.02
DIF22	382	-1.60	1.85	-0.02	0.96 ^{†††}

Note. Valid cases only.

[‡]*p* < .05, two-tailed; ^{†‡}*p* < .01, two-tailed; ^{†††}*p* < .001, two-tailed.

Table D4
Distribution of Overall Importance Battery Items

	<i>f</i>	%	Cumulative %
Most important dimension (IMPMOS) ^a			
Tangibles	14	4	4
Reliability	150	40	44
Responsiveness	98	27	71
Assurance	67	18	89
Empathy	42	11	100
Second most important dimension (IMPSEC) ^b			
Tangibles	14	4	4
Reliability	109	29	33
Responsiveness	139	36	69
Assurance	76	20	89
Empathy	40	11	100
Least important dimension (IMPLEA) ^c			
Tangibles	259	69	69
Reliability	9	2	71
Responsiveness	15	4	75
Assurance	33	9	84
Empathy	60	16	100

Note. Valid cases only.

^a(*N*=371). ^b(*N*=378). ^c(*N*=376).

Table D5

Descriptive Statistics for the Relative Importance Battery Items

	<i>n</i>	<i>M</i>	<i>SD</i>	Skewness	Kurtosis
MIMPTAN	387	14.33	8.01	1.49 ^{†††}	5.19 ^{†††}
MIMPREL	387	24.88	9.97	1.17 ^{†††}	2.92 ^{†††}
MIMPRES	387	22.62	8.30	1.13 ^{†††}	3.78 ^{†††}
MIMPASS	387	21.04	8.81	2.10 ^{†††}	8.30 ^{†††}
MIMPEMP	387	17.13	8.00	1.36 ^{†††}	4.75 ^{†††}

Note. Constant-sum scale, total equals 100. Valid cases only. Cases with a value for at least one scale, and which did not add-up, were made proportional.

^{†††} $p < .001$, two-tailed.

Table D6

Descriptive Statistics for the Satisfaction Items

	<i>n</i>	<i>M</i>	<i>SD</i>	Skewness	Kurtosis
PSAT1 ^a	371	4.72	1.58	-0.20	-0.63 [†]
SAT2	386	4.83	1.49	-0.60 ^{†††}	0.00
PSAT3 ^a	383	6.00	1.63	-1.69 ^{†††}	1.94 ^{†††}
SAT4	375	4.41	1.40	-0.27 [†]	0.13
SAT5	380	4.83	1.57	-0.65 ^{†††}	-0.10

Note. Seven-point Likert scales (1 = *strongly disagree*, 7 = *strongly agree*). Valid cases only.

^aReflected.

[†] $p < .05$, two-tailed; ^{††} $p < .01$, two-tailed; ^{†††} $p < .001$, two-tailed.

APPENDIX E

Means, Standard Deviations, and Correlations for the Data Set Used in Factorial Analysis

	EXP1	EXP2	EXP3	EXP4	EXP5	EXP6	EXP7
EXP1	1.000						
EXP2	0.431	1.000					
EXP3	0.225	0.457	1.000				
EXP4	0.254	0.475	0.448	1.000			
EXP5	0.051	0.114	0.143	0.221	1.000		
EXP6	0.053	0.109	0.272	0.150	0.456	1.000	
EXP7	0.135	0.124	0.240	0.082	0.220	0.487	1.000
EXP8	0.027	0.078	0.238	0.130	0.348	0.552	0.627
EXP9	0.068	0.070	0.194	0.162	0.350	0.504	0.463
EXP10	0.172	0.129	0.238	0.200	0.285	0.421	0.423
EXP11	0.089	0.121	0.279	0.199	0.294	0.518	0.486
EXP12	0.166	0.106	0.295	0.195	0.303	0.532	0.469
EXP13	0.185	0.160	0.261	0.143	0.234	0.427	0.405
EXP14	0.097	0.154	0.148	0.216	0.237	0.290	0.239
EXP15	0.098	0.146	0.161	0.190	0.314	0.456	0.272
EXP16	0.160	0.129	0.282	0.165	0.318	0.463	0.407
EXP17	0.095	0.139	0.209	0.145	0.312	0.466	0.352
EXP18	0.181	0.183	0.171	0.148	0.121	0.302	0.414
EXP19	0.058	0.116	0.139	0.166	0.194	0.335	0.345
EXP20	0.167	0.184	0.245	0.201	0.190	0.316	0.407
EXP21	0.061	0.024	0.241	0.232	0.409	0.537	0.295
EXP22	0.060	0.113	0.329	0.147	0.333	0.520	0.505
PER1	0.265	0.212	0.138	0.101	0.047	0.105	0.085
PER2	0.053	0.192	0.153	0.125	0.003	0.136	0.152
PER3	0.110	0.169	0.142	0.084	0.051	0.144	0.069
PER4	0.025	0.148	0.143	0.321	0.114	0.139	0.041
PER5	0.230	0.168	0.131	0.193	0.006	0.119	0.043
PER6	0.168	0.157	0.122	0.162	-0.006	0.072	0.037
PER7	0.084	0.126	0.087	0.082	0.040	0.100	0.196
PER8	0.097	0.171	0.087	0.144	0.018	0.082	0.107
PER9	0.109	0.183	0.044	0.101	0.062	0.127	0.144
PER10	0.020	0.081	0.048	0.098	-0.047	-0.041	0.027
PER11	0.141	0.124	0.098	0.162	-0.032	-0.025	-0.024
PER12	0.093	0.106	0.098	0.123	0.010	0.002	0.014
PER13	0.101	0.155	0.096	0.096	-0.021	0.024	0.121
PER14	0.137	0.150	0.124	0.182	0.015	0.047	0.066
PER15	0.105	0.128	0.022	0.198	0.012	0.062	-0.022
PER16	0.077	0.105	0.083	0.122	0.049	0.026	0.135
PER17	0.176	0.133	0.040	0.171	-0.012	0.005	0.011
PER18	0.142	0.116	0.050	0.169	-0.047	0.003	0.030
PER19	0.155	0.152	0.088	0.163	0.006	0.043	-0.020
PER20	0.134	0.159	0.113	0.211	0.007	0.048	0.055
PER21	0.157	0.208	0.127	0.249	0.007	0.085	0.056
PER22	0.203	0.164	0.107	0.204	-0.038	-0.009	-0.041
PSAT1	-0.082	-0.028	0.045	0.045	-0.002	0.041	-0.075
SAT2	0.114	0.153	0.105	0.179	-0.059	0.015	-0.014
PSAT3	-0.064	-0.037	-0.043	0.080	0.077	0.083	-0.098
SAT4	0.157	0.147	0.155	0.189	-0.051	0.028	-0.001
SAT5	0.121	0.101	0.117	0.193	-0.073	-0.007	-0.024
\bar{x}	4.43	4.64	5.56	5.42	6.27	6.52	5.94
s	1.62	1.80	1.52	1.53	1.24	0.95	1.24

	EXP8	EXP9	EXP10	EXP11	EXP12	EXP13	EXP14
EXP8	1.000						
EXP9	0.562	1.000					
EXP10	0.557	0.492	1.000				
EXP11	0.561	0.524	0.609	1.000			
EXP12	0.504	0.487	0.445	0.592	1.000		
EXP13	0.470	0.412	0.371	0.433	0.635	1.000	
EXP14	0.252	0.320	0.338	0.263	0.282	0.289	1.000
EXP15	0.352	0.400	0.431	0.366	0.389	0.314	0.531
EXP16	0.472	0.474	0.492	0.434	0.536	0.491	0.449
EXP17	0.388	0.479	0.367	0.426	0.432	0.415	0.446
EXP18	0.231	0.309	0.233	0.256	0.348	0.329	0.386
EXP19	0.364	0.332	0.300	0.209	0.361	0.382	0.258
EXP20	0.330	0.332	0.326	0.310	0.445	0.414	0.359
EXP21	0.433	0.494	0.424	0.415	0.432	0.359	0.407
EXP22	0.475	0.415	0.380	0.393	0.459	0.408	0.363
PER1	0.112	0.166	0.084	0.126	0.180	0.161	0.066
PER2	0.162	0.169	0.068	0.138	0.140	0.140	0.116
PER3	0.088	0.118	0.101	0.116	0.185	0.138	0.203
PER4	0.090	0.098	0.030	0.149	0.069	0.081	0.191
PER5	0.042	0.104	0.086	0.014	0.076	0.052	0.115
PER6	-0.018	0.081	0.070	0.005	0.046	0.006	0.090
PER7	0.109	0.105	0.046	0.014	-0.002	0.019	0.090
PER8	0.055	0.064	0.080	0.062	0.082	0.053	0.117
PER9	0.123	0.142	0.076	0.099	0.099	0.089	0.158
PER10	0.095	0.041	0.079	0.077	0.046	0.012	0.017
PER11	0.023	0.009	0.081	0.106	0.070	0.005	0.084
PER12	0.048	0.076	0.039	-0.012	0.121	0.078	0.112
PER13	0.117	0.057	0.059	-0.007	0.098	0.104	0.131
PER14	0.046	0.064	0.072	0.009	0.072	0.003	0.162
PER15	0.001	0.055	0.093	0.019	0.035	0.014	0.102
PER16	0.129	0.097	0.085	0.071	0.148	0.079	0.192
PER17	0.010	0.065	0.063	0.013	0.132	0.022	0.095
PER18	-0.011	0.052	0.071	0.023	0.102	0.033	0.089
PER19	-0.056	0.022	0.069	0.021	0.018	0.003	0.071
PER20	0.022	0.082	0.057	-0.004	0.085	0.080	0.163
PER21	0.023	0.077	0.082	0.018	0.061	0.047	0.132
PER22	-0.051	0.063	0.066	-0.009	0.054	0.006	0.135
PSAT1	0.023	0.050	0.041	-0.031	-0.035	-0.079	0.066
SAT2	-0.029	0.016	0.015	-0.031	0.017	-0.017	0.014
PSAT3	0.000	0.043	-0.042	-0.031	0.008	-0.068	0.087
SAT4	0.024	0.085	0.024	-0.046	-0.011	0.024	0.002
SAT5	-0.059	0.088	0.025	0.001	0.047	-0.009	0.010
\bar{x}	6.38	6.28	6.34	6.33	6.28	6.05	6.10
s	0.96	1.04	1.03	0.99	1.13	1.31	1.36

	EXP15	EXP16	EXP17	EXP18	EXP19	EXP20	EXP21
EXP15	1.000						
EXP16	0.647	1.000					
EXP17	0.596	0.592	1.000				
EXP18	0.437	0.385	0.487	1.000			
EXP19	0.354	0.342	0.387	0.341	1.000		
EXP20	0.363	0.391	0.373	0.711	0.456	1.000	
EXP21	0.545	0.561	0.547	0.287	0.381	0.374	1.000
EXP22	0.520	0.550	0.575	0.451	0.523	0.526	0.594
PER1	0.116	0.125	0.069	0.114	0.069	0.187	0.106
PER2	0.151	0.092	0.118	0.145	0.094	0.182	0.129
PER3	0.153	0.115	0.171	0.132	0.078	0.147	0.109
PER4	0.185	0.149	0.129	0.182	0.148	0.204	0.189
PER5	0.124	0.052	0.078	0.166	0.045	0.112	0.115
PER6	0.105	0.060	0.090	0.166	0.013	0.092	0.091
PER7	0.056	0.003	0.061	0.173	0.051	0.090	0.044
PER8	0.052	0.008	0.036	0.176	0.040	0.134	0.073
PER9	0.120	0.049	0.053	0.142	0.129	0.157	0.098
PER10	0.016	-0.022	0.002	-0.021	0.025	0.021	0.013
PER11	0.059	0.007	0.013	0.061	-0.002	0.057	0.034
PER12	0.048	0.018	0.083	0.141	0.075	0.187	0.034
PER13	0.071	0.050	0.074	0.157	0.170	0.207	0.039
PER14	0.126	0.049	0.095	0.151	0.032	0.112	0.061
PER15	0.121	0.039	0.056	0.120	0.024	0.111	0.140
PER16	0.113	0.108	0.123	0.183	0.091	0.186	0.104
PER17	0.120	0.038	0.099	0.122	0.054	0.112	0.104
PER18	0.076	0.038	0.071	0.204	0.066	0.188	0.051
PER19	0.137	0.062	0.078	0.030	0.128	0.056	0.085
PER20	0.106	0.073	0.125	0.202	0.126	0.235	0.088
PER21	0.124	0.053	0.049	0.190	0.025	0.131	0.094
PER22	0.124	0.059	0.058	0.123	-0.002	0.106	0.083
PSAT1	0.079	-0.027	-0.052	-0.048	-0.069	0.013	0.131
SAT2	0.049	-0.001	-0.024	0.052	-0.043	0.045	0.040
PSAT3	0.117	0.035	0.030	-0.001	0.020	-0.037	0.130
SAT4	-0.018	-0.005	-0.039	0.030	0.070	0.038	0.023
SAT5	-0.013	-0.022	-0.050	0.067	-0.034	0.060	0.017
\bar{x}	6.42	6.45	5.23	5.78	6.00	5.85	6.52
s	1.04	1.00	1.21	1.37	1.28	1.25	0.93

	EXP22	PER1	PER2	PER3	PER\$	PER%	PER6
EXP22	1.000						
PER1	0.175	1.000					
PER2	0.237	0.579	1.000				
PER3	0.156	0.275	0.460	1.000			
PER4	0.166	0.235	0.362	0.318	1.000		
PER5	0.095	0.299	0.308	0.351	0.306	1.000	
PER6	0.067	0.199	0.298	0.392	0.250	0.713	1.000
PER7	0.141	0.180	0.323	0.325	0.269	0.563	0.674
PER8	0.073	0.236	0.343	0.413	0.302	0.576	0.702
PER9	0.105	0.295	0.298	0.302	0.374	0.515	0.490
PER10	0.053	0.185	0.245	0.360	0.271	0.444	0.494
PER11	0.019	0.209	0.274	0.457	0.285	0.521	0.606
PER12	0.113	0.228	0.301	0.519	0.251	0.519	0.634
PER13	0.153	0.208	0.375	0.486	0.258	0.476	0.573
PER14	0.135	0.269	0.294	0.505	0.256	0.548	0.638
PER15	0.070	0.232	0.273	0.405	0.277	0.576	0.679
PER16	0.104	0.113	0.293	0.474	0.217	0.310	0.400
PER17	0.086	0.245	0.285	0.494	0.214	0.448	0.576
PER18	0.015	0.148	0.273	0.429	0.275	0.456	0.603
PER19	0.114	0.291	0.243	0.231	0.173	0.273	0.344
PER20	0.142	0.218	0.345	0.454	0.313	0.464	0.567
PER21	0.113	0.234	0.269	0.374	0.313	0.595	0.634
PER22	0.095	0.258	0.281	0.437	0.228	0.527	0.641
PSAT1	-0.001	0.165	0.140	0.081	0.136	0.178	0.179
SAT2	-0.017	0.219	0.286	0.347	0.283	0.560	0.656
PSAT3	0.030	0.050	0.086	0.189	0.219	0.218	0.288
SAT4	0.066	0.335	0.341	0.344	0.271	0.468	0.470
SAT5	-0.019	0.251	0.295	0.374	0.272	0.488	0.568
\bar{x}	6.12	4.78	4.81	4.80	5.13	4.54	4.50
s	1.19	1.15	1.30	1.38	1.34	1.50	1.60

	PER7	PER8	PER9	PER10	PER11	PER12	PER13
PER7	1.000						
PER8	0.744	1.000					
PER9	0.606	0.674	1.000				
PER10	0.546	0.640	0.610	1.000			
PER11	0.581	0.685	0.611	0.773	1.000		
PER12	0.586	0.616	0.495	0.632	0.722	1.000	
PER13	0.559	0.608	0.496	0.606	0.665	0.826	1.000
PER14	0.567	0.630	0.577	0.602	0.702	0.806	0.747
PER15	0.575	0.667	0.584	0.590	0.661	0.693	0.644
PER16	0.443	0.506	0.387	0.525	0.553	0.646	0.701
PER17	0.497	0.585	0.505	0.569	0.666	0.714	0.667
PER18	0.545	0.574	0.447	0.515	0.586	0.679	0.650
PER19	0.270	0.314	0.272	0.378	0.399	0.321	0.383
PER20	0.546	0.546	0.459	0.510	0.608	0.741	0.740
PER21	0.607	0.606	0.566	0.472	0.591	0.630	0.586
PER22	0.555	0.595	0.464	0.550	0.639	0.703	0.642
PSAT1	0.166	0.209	0.178	0.159	0.149	0.160	0.147
SAT2	0.580	0.661	0.557	0.565	0.593	0.579	0.558
PSAT3	0.212	0.273	0.217	0.273	0.286	0.309	0.264
SAT4	0.481	0.469	0.421	0.406	0.397	0.499	0.440
SAT5	0.525	0.583	0.455	0.526	0.529	0.577	0.518
\bar{x}	4.31	4.52	4.71	4.64	4.75	4.82	4.61
s	1.54	1.49	1.40	1.63	1.51	1.60	1.60

	PER14	PER15	PER16	PER17	PER18	PER19	PER20
PER14	1.000						
PER15	0.782	1.000					
PER16	0.616	0.565	1.000				
PER17	0.701	0.714	0.687	1.000			
PER18	0.626	0.654	0.639	0.662	1.000		
PER19	0.395	0.432	0.356	0.400	0.390	1.000	
PER20	0.705	0.657	0.641	0.654	0.834	0.523	1.000
PER21	0.683	0.728	0.518	0.617	0.669	0.452	0.702
PER22	0.701	0.690	0.584	0.699	0.715	0.465	0.752
PSAT1	0.178	0.250	0.128	0.195	0.146	0.118	0.135
SAT2	0.603	0.679	0.487	0.599	0.597	0.410	0.596
PSAT3	0.292	0.403	0.272	0.348	0.285	0.229	0.319
SAT4	0.494	0.518	0.351	0.433	0.439	0.340	0.506
SAT5	0.560	0.653	0.456	0.517	0.589	0.371	0.589
\bar{x}	4.85	5.00	5.30	4.99	4.41	4.87	4.47
s	1.59	1.47	1.50	1.46	1.59	1.72	1.60

	PER21	PER22	PSAT1	SAT2	PSAT3	SAT4	SAT5
PER21	1.000						
PER22	0.762	1.000					
PSAT1	0.229	0.186	1.000				
SAT2	0.631	0.598	0.309	1.000			
PSAT3	0.356	0.343	0.329	0.406	1.000		
SAT4	0.515	0.502	0.224	0.615	0.259	1.000	
SAT5	0.602	0.603	0.278	0.757	0.395	0.705	1.00
\bar{x}	4.63	4.51	4.73	4.85	6.00	4.41	4.83
s	1.53	1.53	1.55	1.48	1.63	1.39	1.57

APPENDIX F

EQS Syntax for the Estimation of the Simultaneous Model

/TITLE

2 second-order quality factors+satisfaction,structural,correlated errors

Variant=theory

/SPECIFICATIONS

DATA='Questi03.ess'; VARIABLES=187; CASES= 384;

METHODS=ML,Robust;

MATRIX=RAW;

/LABELS

V1=EXP1; V2=EXP2; V3=EXP3; V4=EXP4; V5=EXP5;
 V6=EXP6; V7=EXP7; V8=EXP8; V9=EXP9; V10=EXP10;
 V11=EXP11; V12=EXP12; V13=EXP13; V14=EXP14; V15=EXP15;
 V16=EXP16; V17=EXP17; V18=EXP18; V19=EXP19; V20=EXP20;
 V21=EXP21; V22=EXP22; V23=PER1; V24=PER2; V25=PER3;
 V26=PER4; V27=PER5; V28=PER6; V29=PER7; V30=PER8;
 V31=PER9; V32=PER10; V33=PER11; V34=PER12; V35=PER13;
 V36=PER14; V37=PER15; V38=PER16; V39=PER17; V40=PER18;
 V41=PER19; V42=PER20; V43=PER21; V44=PER22; V45=PSAT1;
 V46=SAT2; V47=PSAT3; V48=SAT4; V49=SAT5; V50=DIF1;
 V51=DIF2; V52=DIF3; V53=DIF4; V54=DIF5; V55=DIF6;
 V56=DIF7; V57=DIF8; V58=DIF9; V59=DIF10; V60=DIF11;
 V61=DIF12; V62=DIF13; V63=DIF14; V64=DIF15; V65=DIF16;
 V66=DIF17; V67=DIF18; V68=DIF19; V69=DIF20; V70=DIF21;
 V71=DIF22; V72=EMPL; V73=GEND; V74=EDUC; V75=STAB;
 V76=OWNE; V77=INCO; V78=AGEG; V79=LANG; V80=CASE;
 V81=MIMPTAN; V82=MIMPREL; V83=MIMPRES; V84=MIMPASS; V85=MIMPEMP;
 V86=TOTMIMP; V87=EXPTAN; V88=EXPREL; V89=EXPRES; V90=EXPASS;
 V91=EXPEMP; V92=PERTAN; V93=PERREL; V94=PERRES; V95=PERASS;
 V96=PEREMP; V97=DIFTAN; V98=DIFREL; V99=DIFRES; V100=DIFASS;
 V101=DIFEMP; V102=WEITAN; V103=WEIREL; V104=WEIRES; V105=WEIASS;
 V106=WEIEMP; V107=WEXP1; V108=WEXP2; V109=WEXP3; V110=WEXP4;
 V111=WEXP5; V112=WEXP6; V113=WEXP7; V114=WEXP8; V115=WEXP9;
 V116=WEXP10; V117=WEXP11; V118=WEXP12; V119=WEXP13; V120=WEXP14;
 V121=WEXP15; V122=WEXP16; V123=WEXP17; V124=WEXP18; V125=WEXP19;
 V126=WEXP20; V127=WEXP21; V128=WEXP22; V129=WPER1; V130=WPER2;
 V131=WPER3; V132=WPER4; V133=WPER5; V134=WPER6; V135=WPER7;

V136=WPER8; V137=WPER9; V138=WPER10; V139=WPER11; V140=WPER12;
 V141=WPER13; V142=WPER14; V143=WPER15; V144=WPER16; V145=WPER17;
 V146=WPER18; V147=WPER19; V148=WPER20; V149=WPER21; V150=WPER22;
 V151=WDIF1; V152=WDIF2; V153=WDIF3; V154=WDIF4; V155=WDIF5;
 V156=WDIF6; V157=WDIF7; V158=WDIF8; V159=WDIF9; V160=WDIF10;
 V161=WDIF11; V162=WDIF12; V163=WDIF13; V164=WDIF14; V165=WDIF15;
 V166=WDIF16; V167=WDIF17; V168=WDIF18; V169=WDIF19; V170=WDIF20;
 V171=WDIF21; V172=WDIF22; V173=WEXPTAN; V174=WEXPREL; V175=WEXPRES;
 V176=WEXPASS; V177=WEXPEMP; V178=WPERTAN; V179=WPERREL; V180=WPERRES;
 V181=WPERASS; V182=WPEREMP; V183=WDIFTAN; V184=WDIFREL; V185=WDIFRES;
 V186=WDIFASS; V187=WDIFEMP;

/EQUATIONS

V1 = + 1.0F1 + E1;
 V2 = + *F1 + E2;
 V3 = + *F1 + E3;
 V4 = + *F1 + E4;
 V5 = + 1.0F2 + E5;
 V6 = + *F2 + E6;
 V7 = + *F2 + E7;
 V8 = + *F2 + E8;
 V9 = + *F2 + E9;
 V10 = + 1.0F3 + E10;
 V11 = + *F3 + E11;
 V12 = + *F3 + E12;
 V13 = + *F3 + E13;
 V14 = + 1.0F4 + E14;
 V15 = + *F4 + E15;
 V16 = + *F4 + E16;
 V17 = + *F4 + E17;
 V18 = + 1.0F5 + E18;
 V19 = + *F5 + E19;
 V20 = + *F5 + E20;
 V21 = + *F5 + E21;
 V22 = + *F5 + E22;
 V23 = + 1.0F6 + E23;
 V24 = + *F6 + E24;
 V25 = + *F6 + E25;


```

V26 = + *F6 + E26;
V27 = + 1.0F7 + E27;
V28 = + *F7 + E28;
V29 = + *F7 + E29;
V30 = + *F7 + E30;
V31 = + *F7 + E31;
V32 = + 1.0F8 + E32;
V33 = + *F8 + E33;
V34 = + *F8 + E34;
V35 = + *F8 + E35;
V36 = + 1.0F9 + E36;
V37 = + *F9 + E37;
V38 = + *F9 + E38;
V39 = + *F9 + E39;
V40 = + 1.0F10 + E40;
V41 = + *F10 + E41;
V42 = + *F10 + E42;
V43 = + *F10 + E43;
V44 = + *F10 + E44;
V45 = + 1.0F11 + E45;
V46 = + *F11 + E46;
V47 = + *F11 + E47;
V48 = + *F11 + E48;
V49 = + *F11 + E49;
F1 = + 1.0F20 + D1;
F2 = + *F20 + D2;
F3 = + *F20 + D3;
F4 = + *F20 + D4;
F5 = + *F20 + D5;
F6 = + 1.0F30 + D6;
F7 = + *F30 + D7;
F8 = + *F30 + D8;
F9 = + *F30 + D9;
F10 = + *F30 + D10;
F30 = + *F20 + D30;
F11 = + *F20 + *F30 + D11;
/VARIANCES

```

F20 = *;
E1 = *;
E2 = *;
E3 = *;
E4 = *;
E5 = *;
E6 = *;
E7 = *;
E8 = *;
E9 = *;
E10 = *;
E11 = *;
E12 = *;
E13 = *;
E14 = *;
E15 = *;
E16 = *;
E17 = *;
E18 = *;
E19 = *;
E20 = *;
E21 = *;
E22 = *;
E23 = *;
E24 = *;
E25 = *;
E26 = *;
E27 = *;
E28 = *;
E29 = *;
E30 = *;
E31 = *;
E32 = *;
E33 = *;
E34 = *;
E35 = *;
E36 = *;

```
E37 = *;  
E38 = *;  
E39 = *;  
E40 = *;  
E41 = *;  
E42 = *;  
E43 = *;  
E44 = *;  
E45 = *;  
E46 = *;  
E47 = *;  
E48 = *;  
E49 = *;  
D1 = *;  
D2 = *;  
D3 = *;  
D4 = *;  
D5 = *;  
D6 = *;  
D7 = *;  
D8 = *;  
D9 = *;  
D10 = *;  
D11 = *;  
D30 = *;  
/COVARIANCES  
E1,E23 = *;  
E2,E24 = *;  
E3,E25 = *;  
E4,E26 = *;  
E5,E27 = *;  
E6,E28 = *;  
E7,E29 = *;  
E8,E30 = *;  
E9,E31 = *;  
E10,E32 = *;  
E11,E33 = *;
```

```
E12,E34 = *;  
E13,E35 = *;  
E14,E36 = *;  
E15,E37 = *;  
E16,E38 = *;  
E17,E39 = *;  
E18,E40 = *;  
E19,E41 = *;  
E20,E42 = *;  
E21,E43 = *;  
E22,E44 = *;  
/CONSTRAINTS  
/LMTEST  
PROCESS=SIMULTANEOUS;  
SET=PVV,PFV,PFF,PDD,GVV,GVF,GFV,GFF,BVF,BFF;  
/WTEST  
PVAL=0.05;  
PRIORITY=ZERO;  
/PRINT  
digit=3;  
linesize =80;  
fit=all;  
/END
```

APPENDIX G

Estimates of the Standardized Factor Loadings and Statistics for the Simultaneous Model

Table G1

Estimates of the Standardized Factor Loadings and Statistics for the Expectation Variables in the Simultaneous Model

	Tangibles	Reliability	Responsiveness	Assurance	Empathy	SMC
EXP1	.45 ^a					.20
EXP2	.73					.53
EXP3	.65					.42
EXP4	.64					.41
EXP5		.48 ^a				.23
EXP6		.74				.55
EXP7		.69				.47
EXP8		.77				.59
EXP9		.71				.50
EXP10			.68 ^a			.46
EXP11			.75			.57
EXP12			.78			.61
EXP13			.67			.45
EXP14				.59 ^a		.35
EXP15				.78		.61
EXP16				.81		.66
EXP17				.76		.57
EXP18					.62 ^a	.39
EXP19					.60	.36
EXP20					.69	.47
EXP21					.68	.47
EXP22					.81	.65
γ_i	.37 ^a	.93	.92	.84	.87	
R^2	.14	.87	.85	.71	.75	

Note. SMC = squared multiple correlation.

^aParameter fixed during estimation to assign a metric to the scale.

Table G2

Estimates of the Standardized Factor Loadings and Statistics for the Perception Variables in the Simultaneous Model

	Tangibles	Reliability	Responsiveness	Assurance	Empathy	SMC
PER1	.62 ^a					.38
PER2	.80					.64
PER3	.61					.37
PER4	.46					.21
PER5		.73 ^a				.53
PER6		.83				.69
PER7		.82				.67
PER8		.87				.76
PER9		.72				.51
PER10			.75 ^a			.56
PER11			.83			.69
PER12			.90			.82
PER13			.87			.75
PER14				.88 ^a		.77
PER15				.86		.74
PER16				.73		.53
PER17				.82		.68
PER18					.85 ^a	.72
PER19					.54	.28
PER20					.89	.80
PER21					.83	.68
PER22					.87	.75
γ_i	.56 ^a	.88	.95	.98	.92	
R^2	.31	.77	.91	.97	.85	

Note. SMC = squared multiple correlation.

^aParameter fixed during estimation to assign a metric to the scale.

Table G3

Estimates of the Standardized Factor Loadings and Statistics for the Satisfaction Variables in the Simultaneous Model

	Satisfaction	SMC
PSAT1	.33 ^a	.11
SAT2	.87	.76
PSAT3	.45	.20
SAT4	.74	.55
SAT5	.88	.77

Note. SMC = squared multiple correlation.

^aParameter fixed during estimation to assign a metric to the scale.

APPENDIX H

Estimates of the Standardized Factor Loadings and Statistics for the Perception-Only Model

	Tangibles	Reliability	Responsiveness	Assurance	Empathy	Satisfaction
PER1	.61 ^a					
PER2	.78					
PER3	.63					
PER4	.47					
PER5		.72 ^a				
PER6		.83				
PER7		.82				
PER8		.87				
PER9		.72				
PER10			.75 ^a			
PER11			.83			
PER12			.91			
PER13			.86			
PER14				.88 ^a		
PER15				.86		
PER16				.73		
PER17				.83		
PER18					.86 ^a	
PER19					.53	
PER20					.90	
PER21					.82	
PER22					.87	
PSAT1						.33 ^a
SAT2						.87
PSAT3						.45
SAT4						.74
SAT5						.88
γ_i	.59 ^a	.87	.95	.98	.92	
R^2	.35	.76	.91	.97	.85	

^aParameter fixed during estimation to assign a metric to the scale.

APPENDIX I

Estimates of the Standardized Factor Loadings for the Unifactorial Models

	Tangibles	Reliability	Responsiveness	Assurance	Empathy	Satisfaction
PER1	.62 ^a					
PER2	.92					
PER3	.51					
PER4	.41					
PER5		.72 ^a				
PER6		.82				
PER7		.83				
PER8		.88				
PER9		.71				
PER10			.75 ^a			
PER11			.82			
PER12			.91			
PER13			.87			
PER14				.87 ^a		
PER15				.86		
PER16				.72		
PER17				.84		
PER18					.87 ^a	
PER19					.53	
PER20					.92	
PER21					.80	
PER22					.84	
PSAT1						.33 ^a
SAT2						.83
PSAT3						.44
SAT4						.75
SAT5						.92

^aParameter fixed during estimation to assign a metric to the scale.