Statistica & Applicazioni - Vol. IX, n. 1, 2011, pp. 79-92

THE CUSTOMER SATISFACTION MEASURE IN PUBLIC SERVICES: THE CASE OF AN ITALIAN CHAMBER OF COMMERCE

Laura Pagani*
Gian Pietro Zaccomer*
Maria Chiara Zanarotti**

SUMMARY

Nowadays, in Italy, surveys carried out to measure Customer Satisfaction in Public Services are going to become more important and systematic. The new directive form Public Function Office (2004) emphasizes the fundamental role of users' opinions to drive and possibly improve public services. The administrators have to consider users' opinions, because these judgments are useful to offer more acceptable and appreciable services. Measure Customer Satisfaction in Public services is not a simple task and for this reason it is at the center of growing attentions. In this paper, after dealing some theoretical aspects related to this matter, the case of an Italian Chamber of Commerce is considered and different statistical methods to analyze users' satisfaction data are proposed.

Keywords: *Service Quality, Heterogeneity and Dissimilarity Index, Rasch Analysis, Overall Satisfaction.*

1. INTRODUCTION

The absolute necessity of satisfying the even more demanding customers is a well-known phenomena in the domestic and world markets. Also in the public sector we can observe the growth and the great differentiation of users' needs (not only citizens, but also enterprises and other private corporations). Therefore, in the public sector, such necessity is closely connected to the renewing process of the organization, to the way of work and to the economical and financial management of the *Public Administration* (PA).

In Italy, as in many other European Countries, this process has not been completed yet. The main problem is that the process cannot start without a deep cultural change based on the *evaluation of ''enterprises' performances''*.

This will allow PA to come out from their self-referral, which is an attitude that has always placed the users on a different level with respect to the service supplier.

* Dipartimento di Scienze Economiche e Statistiche - Università di Udine - via Treppo, 18 - 33100 UDINE (e-mail: pagani@uniud.it; zaccomer@uniud.it).
** Dipartimento di Scienze Statistiche - Università Cattolica del Sacro Cuore - Via Necchi, 9 - 20123 MILANO (e-mail: ✉ chiara.zanarotti@unicatt.it).

Though L. Pagani wrote Sections 4 and 5, G.P. Zaccomer wrote Sections 1 and 2, M.C. Zanarotti wrote Section 3, 4 and 5, the work reflects the common thinking of the authors.

In this work, the enterprise's performance that we consider is the *quality of a public service*. If we refer to the classical definition, this quality is the whole characteristics of a public service that are necessary to satisfy the users' needs, so we can see the closely relation between the quality level of service and the *Customer Satisfaction* (CS). Moreover, this implies the need of defining the quality of the service through a statistical method able to return the most possible real measure. In fact, also the Italian Department for the Public Function underlines the danger of a mismanage CS survey which ''risks to become an action facade instead a real tool of design and improvement of services'' (Tanese, Negro and Gramigna, 2003, p.13).

Section 2 is devoted to describe the underlying Italian framework about CS in PA and introduce specific problems in dealing with CS in Italian Chamber of Commerce. In Section 3 methodological tools the authors suggest to use are summarized and main results from data analysis is performed in Section 4. Concluding remarks are outlined in Section 5.

2. THE UNDERLYING ITALIAN FRAMEWORK

The CS definition comes from the mass production crisis, when the marketing approach became a real necessity for enterprises' survival.

Since the 80s, first in Japan and then in the occidental world, the customer oriented management leaded to the formalization of the *Total Quality* (TQ) and to a model defined *Total Quality Management* (TQM). In the TQM model, the evaluation of quality with a scientific approach is an inexorable necessity, and, therefore, also the CS evaluation. Hence, the total quality, the service quality and the CS are closely connected.

As we said before, the public sector has gradually adopted the marketing tools, adapting them to its own characteristics but, in this development, the most important point regards the role of the marketing function inside its decision processes (Kotler, 1978). As for an enterprise the marketing approach is a necessity, for a PA it represents a key of evolution that allows the passage from a bureaucratic oriented system to a more modern *user oriented* one.

Informally, we can define the CS of a PA as a complex procedure which uses many different statistic tools to collect and process information for the continuous improvement of the quality of services. The most important step of this procedure regards the evaluation process with a statistical survey, called *CS survey*. Usually, this is a self-administered survey realized by using a questionnaire. The CS survey, that involves the PA users, allows to obtain the necessary information to manage the organization on each level: from the lower level (*i.e.* not clear documents, the lacking in skills of the employees) to the higher level, which can also lead to the reengineering of the service supplying process. Moreover, the information which are obtained from the CS survey are used both for the internal and external communication of the administration. The internal communication aims to transmit the new direction of the administration to the employees; the external communication is carry out to let the

citizen know the results obtained, as established by the principles of transparency which have been recently adopted by the Italian public administration. Therefore, the PA's CS survey should be a real ''*listening method*'' for the users, but this can be completely realized only if the administration uses the obtained results to improve its services.

These are the fundamental theoretical principles. Recently, also the Italian Public Administration has introduced new and innovative rules for the CS. First of all the Directive of the Ministry of Public Function dated March 24th 2004 (G.U. nr. 80 of April 5th 2004) that is a fundamental step because single public administrations must implement most activities, as the engineering and the carrying out of periodical CS surveys, the diffusion of results and their employing by the government and direction bodies. Finally, that document also recalls the role of Public Relation Offices, which are designated by the L. 150/2000, art. 8, to verify the quality of services supplied, by listening to users and also through internal communication. At a ministerial level they recognized the implicit difficulty of the administrations to undertake this path because they often didn't have (and don't still have) the skills to design and manage CS surveys. To help these administrations to better respond to the challenges proposed by this Directive, a specific manual was realized (in bibliography referred as (Tanese *et al.*, 2003). From a statistical point of view, the ministerial manual of 2003 is only the starting point. It highlights the basic aspects of the organization and the collecting data, but it does not closely examine the methods used to build statistic models, while it treats customers' appraisals as quantitative variables.

Subsequently, the Ministry of Innovation and New Technologies issued a Directive of July 27th 2005, according with the Ministry of Public Function, which defined the quality of on-line services and the evaluation of CS. This last Directive regards most of all services which are not supplied by a traditional front-line office (web, digital tv, mobile phones, etc.). It is very important because it confirms the ministerial direction and strongly underlines the importance of the CS surveys which must be periodically developed with appropriate methods.

Recently, the Ministry of Reform and Innovation in the public administration has issued a new Directive (December, 19th 2006) which formally recalls the administrations to the concept of quality, continuous improvement and self-evaluation, by quoting the CS surveys. It also recognizes that their use in the Italian PA is still insufficient, incomplete and occasional, also with respect to the involvement of the personnel and its training. With respect to these Directives, nowadays many public administrations have not complied yet, that is they have not formally set up their CS surveys. Therefore, to accelerate the process modernization of the PA, the Ministry of Public Administration and Innovation is studying new rules for 2010 to force the adoption of some technologic innovations, as the certificate e-mail, and the evaluation system of CS.

The present study deal with the case of CS survey in Italian Chamber of Commerce (CC). More precisely, data analysis performed in Section 4 refer to the CC of Udine. This Chamber has developed the survey every year since 1997 by the cooperation between its Public Relation Office and some external collaborators.

The survey involves the front-line offices, but also those which have an on-line communication with their users as defined by the ministerial directives. For this reason, different offices diversify the items of their questionnaires, but also the methods to collect data. Zaccomer and Marton (2009) already deal with CC data for the years 2006 and 2007. The study for the years 2006 and 2007 is our starting point, but in the next sections we would try to give an answer to new operative exigency, that is the possibility to obtain a measure of the global satisfaction with these kind of dataset. The possibility to obtain a measure of the global satisfaction is essential for the decisional maker of the CC, because, having repeated the analysis in different years, they should immediately understand if there is a progressively improvement, as ministerial directives request. Nevertheless, this important necessity hides another big problem: is it possible to compare any offices and report the evaluation differences between them? In fact, with the analysis developed by Zaccomer and Marton (2009) this comparisons are not possible.

3. METHODS OF ANALYSIS

Literature on statistical models and methods to measure CS has growth enormously in last decades, at the same time CS became a strategic matter in marketing and business research. In particular, the methodological problem arises when satisfaction regards services and not tangible consumer goods. Satisfaction is strictly linked to quality, and if quality regards services, no objective characteristics of good can be measured (for example, durability, presence/absence of defects, and so on), so it is necessary to refer to an abstract construct to measure service quality. ''Service quality is an elusive and abstract construct that is difficult to define and measure'' (Cronin and Taylor, 1992). Many PA (also the CC considered in this work), for example, make use of a simplified version of the so called SERVQUAL model (Parasuraman, Zeithaml and Berry, 1988) to analyze customers data: in particular, by using an evaluation ten-point scale, the Customer Satisfaction Index of each structure is computed as an arithmetic mean of the perceived quality weighted by the expected quality. This method follows the principles of the manual released by the Ministry of the Public Function, because, even if it is not clearly quoted, the model is based on a mean with the hypothesis that the customers' valuations are quantitative.

In this work we propose to use different tools specifically driven by targets we want to pursue. Data considered in the next sections come, as already stated, from different offices of a CC and are collected using different questionnaires that are formed by some common items and by unique items. Responses to each item are collected by an ordinal scale, more precisely by scores ranking from one (very negative valuation) to ten (very positive valuation). No global satisfaction item is present in questionnaires. The goal will be to manage data to obtain information regarding not only single offices to improve service for less satisfactory aspects but also regarding the whole agency to compare office performances and to monitor the whole agency during the time. In this paper we consider and compare three different methods of

analysis that can be applied to the different objectives required. Firstly an explorative data analysis is performed, separately office by office, by ranking items to see if their rates are similar between offices and to get an idea about the effective use of the scale. Then a more sophisticated analysis is performed using Rasch Model. The aim is to analyzed data all together (*i.e.* the whole agency as a unique) as well as to perform separate analysis (office by office). As outlined in the sequel, this analysis should allow to calibrate the questionnaires and to obtain customer satisfaction measures for offices comparison. Finally, to obtain an overall satisfaction measure, an index of comparative performance based on dissimilarity measure is proposed.

As already stated, to conduct a preliminary explorative analysis separately office by office, we use a method of scoring. Specifically, we use a method of items ranking based on median and heterogeneity index (HI) similar to the method so called *scorecard* suggested by Giudici (see, among others, Giudici, 2007). The method of ranking we consider in this paper works in two step: at the first step the median is computed for each item of the questionnaires; items are then ordered from the one with the highest median to the one with the lowest median. Notice that some items can share the same median, so items can result only partially ordered. At second step, an index of heterogeneity is calculated, and items with the same median are ordered using the HI: at the first place we put item with the lowest value of HI, second position is for item with second less great value of HI, and so on. After consider items with higher median, the procedure jumps to items with the second value in the median order and a complete ordering of these items is then obtain using again HI criterion. The procedure ends when all items are ordered and a full ranking is obtained. The heterogeneity index adopted in this procedure (Rampichini, Grilli and Petrucci, 2000) is one specifically devoted for ordinal data and it is denoted with d* in Leti (1983). This HI is developed considering cumulative and retro-cumulative (relative) frequencies and take value zero for minimum heterogeneity and value one (or very closed to one, if n is odd and not very large) for maximum heterogeneity. Maximum heterogeneity is achieved when observations are equally shared between the first and the last response categories (*i.e.* the lowest and the higher). Indicating with $k=1,\ldots, K$ the K ordinate response categories, index d* take the form:

$$\mathrm{d}^* = \frac{4\sum_{k=1}^{K-1} F_k(1-F_k)}{K-1} \tag{1}$$

where F_k are cumulative relative frequencies.

The above ranking procedure (median-HI criterion) will be applied in Section 5 separately office by office to highlight aspects that are more problematic and are first to be improved.

A second tool we consider to carry out the analysis is the so called *Rasch Model* (RM). RM (Rasch, 1960), originally introduced in psychometrics field, is a latent structure model for the analysis of item response data nowadays used in many different contexts, and recently introduced also in CS data analysis (for example, among

others, Bond and Fox, 2001; Bertoli-Barsotti, 1999; Pagani and Zanarotti, 2010). Through RM it is possible to measure CS (the so called *latent* treat) from answers given by n users to J items, each with K ordinal categorical responses. This measure is based on the hypothesis that probability of the answer each user gives to each item devoted to collect opinions about quality is a function of three sets of parameters. The first is the set of *Person Location Parameters* (PLP, denoted θ_i). These parameters express users' satisfaction and reflect all individual and context elements that can influence satisfaction. The second set is the one of *Item Location Parameters* (ILP, denoted β_j). These parameters denote qualitative level embedded in each facet of the service indicated by each item. The last set of parameters is the *Thresholds parameters* set (denoted τ_k). These parameters are related to consecutive couples of response categories and limit respective intervals. Threshold parameters represent difficulty to endorse one response category instead of the previous one. This last set of parameters is due to the fact that each item has more than two response categories, because responses in CS survey are usually collected using ordinal categorical scale. So the Rasch model considered is a polytomous one, where it is supposed that respondents give a judgement to each item using an ordinal categorical scale. The polytomous Rasch model considered in this paper is the so called *Rating Scale Model* (RSM), where the *logit* of consecutive response categories (k, $k-1$) is:

$$\ln \frac{P(X_{ij}=k)}{P(X_{ij}=k-1)} = \theta_i - (\beta_j + \tau_k) \qquad \text{were: } \sum_{k=1}^{K} \tau_k = 0 \tag{2}$$

being X_{ij} the response of person i $(i = 1, ..., n)$ to item j $(j = 1, ..., J)$.

Looking at expression (2), it is clear that RSM supposes that probability user *i-th* gives answer k instead $k-1$ to item *j-th* is bigger if his satisfaction level is higher than quality level expressed by category *k-th*. Model (2) allows to associate a measure (the estimate values of the parameters) on an interval scale to both users and items: through this measure users and items can be located univocally on the same quantitative *continuum*.

Some assumptions are fundamental in RM (*unidimensionality* and *local independence*, for example) for parameters identifiability and estimation: a fully examination of these assumptions is beyond the scope of this paper. For a detailed discussion see, for example, Fischer and Molenaar (1995).

The first application of the RM to dataset of CC of Udine was proposed in Zaccomer and Marton (2009). The authors, by using a different formulation of RM - the so called Partial Credit Model (PCM) - highlighted some differences in results obtained using PCM with respect to the SERVQUAL approach. These differences are consequence of the theoretical ones that are the foundations of the two methods: for example, RM only considers the perceived quality and does not use the expected one as system of weights used, on the contrary, in SERVQUAL. In fact, the last model computes the aggregate Customer Satisfaction Index by a weighted mean of the total score of each item from all users, and these weights are not used in Rasch model (for more details, see Zaccomer and Marton, 2009).

The RSM will be used in Section 4 for separate analysis office by office as well as for the analysis of the whole agency.

The last method of analysis we consider is based on *Dissimilarity Index* (DI) between ordinal distribution. The idea is that observed distribution of responses can be compared with a theoretical one, selected as comparative model. Capursi and Porcu (2001) suggested to use as theoretical distribution the optimal one, so called because it is the distribution were all subjects are suppose to choose the higher response category, *i.e.* all users give very positive evaluation (mark ten, in our application). The DI we use in this paper is the *simple relative dissimilarity index* (see, for example, Leti, 1983), denoted with z*:

$$z^* = \frac{1}{K-1}\sum_{k=1}^{K-1}\left|F_k^O - F_k^T\right| \tag{3}$$

Where: F_k^O is the observed cumulative distribution and F_k^T is the theoretical one. Since in the last distribution F_k^T is zero for $k = 1, ..., K-1$ and one for $k = K$, index z* in this case is equal to:

$$z^* = \frac{1}{K-1}\sum_{k=1}^{K-1}F_k^O \tag{4}$$

Capursi and Porcu (2001) suggest to built a *comparative performance indicator* (CPI) by adding the complement to one of index z* calculated for each item and then to divide this sum by its maximum that is equal to *J*. The *comparative performance indicator* is given by:

$$CPI_h = \frac{1}{J}\sum_{j=1}^{J}(1 - z_j^*) \tag{5}$$

Where h indicates, in our application, the h^{th} office ($h = 1, ..., H$). Since index z* given in (4) assumes value zero when observed distribution is similar to the theoretical one and this represent the situation of maximum satisfaction, CPI_h given with (5) considers complement to one of each z_j*, so that CPI_h increases as satisfaction increases and decreases as satisfaction decreases. When observed distributions are all equal to the optimal one (all judgements are on the highest response category) all the addendum in (5) are 1 and their summation is equal to J . Then, by dividing the sum for *J*, every CPI_h takes values from zero (minimum satisfaction) to one (maximum satisfaction).

Through these CPI_h indexes it is possible to perform comparisons between offices, since the absence of global satisfaction item in questionnaires does not allow other prompt tools for comparison. Finally, a global satisfaction index (GSI, taking again values between zero and one) is obtained as an arithmetic mean of CPI_h, i.e.:

$$GSI = \frac{1}{H}\sum_{h=1}^{H}CPI_h \tag{6}$$

4. EMPIRICAL RESULTS

The data considered in this paper are responses to questionnaires administered by the CC of Udine to its users during the last two months of 2007 and refer to five front-line offices. The structures involved were: Agricultural Office, Enterprises Development Office, Gasoline Office, Internationalization Office and Registry Service Office. For privacy reasons the five offices are labelled with capital letters from A to E.

Since questionnaires of the five offices are partially different, the resulting design is the so called *common-item nonequivalent groups design* (Kolen and Brennan, 2004), in which different forms of a questionnaire share a set of items in common and different groups of respondents are administrated the different forms. The number and the structure of the items forming questionnaires are reported in Table 1.

TABLE 1. - *Structure of questionnaires and number of valid responses*

Request: *Mark each item from 1 (very negative valuation) to 10 (very positive valuation)*		**Office**				
Item		**A**	**B**	**C**	**D**	**E**
I01	Clarity and completeness of the information received	x	x	x	x	x
I02	Competence of operator at the counter	x	x		x	x
I03	Operator courtesy (and availability)	x	x	x	x	x
I04	Length and timetable of the service	x	x	x		
I05	Availability of the service by e-mail or by internet	x				
I06	Availability of clear and complete information (in office or website)	x			x	
I07	Speed in waiting time at the counter		x	x		
I08	Operator speed in the delivery of the service at the counter		x	x		
I09	Visibility and clearness of notice information		x			
I10	Availability and functionality of waiting rooms			x		
I11	Clearness and simplicity in forms compilation			x		x
I12	Easy access to the office by phone				x	x
I13	Reliability and accuracy of the delivered service				x	x
I14	Availability of information about contribution on office website					x
Number of items for each office		**6**	**7**	**7**	**6**	**7**
Number of valid questionnaires for each office		**90**	**112**	**109**	**112**	**132**

For example, the 90 users from office A were faced to items from one to six, the 112 users from office B were faced to items from one to four and from seven to nine, and so on. Response categories are the same for all administrated questionnaires: a ten-point ordinal response scale (from 1 = *very negative valuation* to 10 = *very posi-*

tive valuation). The total number of items is 14, with four unique items (I05, I09, I10 and I14), six items common to two questionnaires (I06, I07, I08, I11, I12 and I13), one item common to three (I04), one item common to four (I02) and two items common to all questionnaires (I01 and I03).

The first step of the analysis considers ranking procedure with median-HI criterion (explained in Section 3), office by office. Results are reported in Table 2. Observing the order, we can see on the left the best aspect (first item) and on the right the worst aspect (last item) for each office.

TABLE 2. - *Items ranking using median-HI criterion*

Office	Item ranking (from better to worst one)						
	Best	←	←		→	→	Worst
A	**I03** (10; 0.119)	**I02** (10; 0.152)	**I01** (10; 0.168)	**I05** (10; 0.235)	**I06** (10; 0.321)	**I04** (9; 0.365)	
B	**I03** (9; 0.285)	**I02** (9; 0.297)	**I08** (9; 0.313)	**I01** (9; 0.315)	**I07** (8; 0.359)	**I04** (8; 0.390)	**I09** (8; 0.416)
C	**I10** (8; 0.402)	**I11** (8; 0.422)	**I03** (8; 0.425)	**I08** (8; 0.426)	**I01** (8; 0.439)	**I07** (8; 0.448)	**I04** (8; 0.544)
D	**I03** (9; 0.289)	**I12** (9; 0.305)	**I02** (8; 0.291)	**I06** (8; 0.301)	**I01** (8; 0.322)	**I13** (8; 0.342)	
E	**I03** (9; 0.260)	**I02** (9; 0.283)	**I01** (9; 0.344)	**I13** (9; 0.381)	**I12** (9; 0.393)	**I11** (8; 0.349)	**I14** (8; 0.354)

Note: in brackets median and HI values.

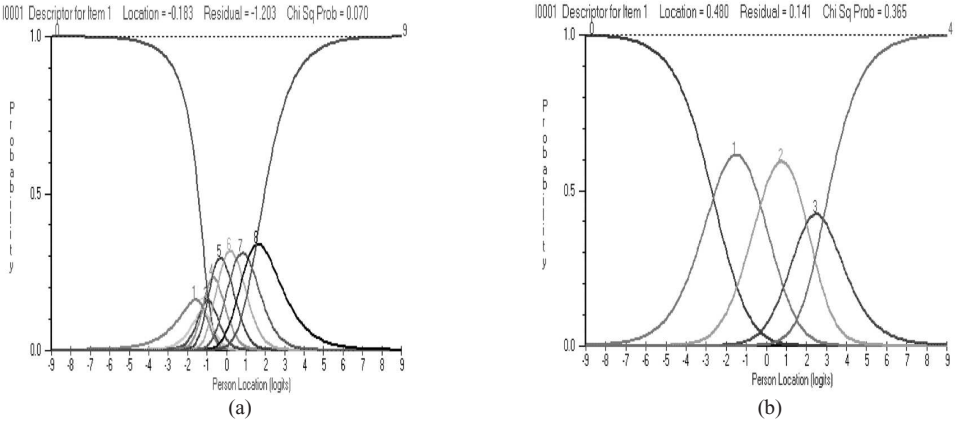
For example, for office A, users are generally very satisfied (all median are equal to ten), more for operator courtesy and competence, less for availability of clear and complete information and for length and timetable of the service. Except for office C, items I03 and I02 (operator courtesy and competence) are, if present, in first positions. *Vice versa*, item I04 (length and timetable of the service), if present, tends to be situated at the end of the order.

The second step consists in using the RSM separately for each office in order to obtain more information about the quality of the service and to check the validity of the ordinal 10 point scale, office by office. Questionnaires are analysed separately with RUMM 2010 (Andrich, Sheridan and Luo, 2004), a standard software for RM. The first analysis of the data reveals that the ten-point ordinal scale is not adequate: categories from 1 to 5 (representing the set of ''unsatisfied'' responses) are too rare for every office, with very low or equal to zero frequencies and, consequently, thresholds are disordered (see, for example, Figure 1). The problem is easily solved collapsing the redundant categories into adjacent categories (Bond and Fox, 2001, Zanarotti and Pagani, 2003).

After several collapsing category procedures a new appropriate ordinal five-point scale is found: 1 = *negative valuation* (marks from 1 to 5), 2 = *sufficient valuation* (marks 6 and 7) and 3 = *good valuation* (mark 8), 4 = *very good valuation* (mark 9) and 5 = *optimum valuation* (mark 10). Figure 1 shows an example of Category Prob-

ability Curve (CPC) for the original ten-point scale, with disordered thresholds (a) and for the final five-point scale (b) with ordered thresholds. In CPC the probability of responding to one of the categories is function of the PLP (logits).

FIGURE 1. - *CPC item I01, office D before (a) and after (b) collapsing categories*



(a) (b)

Finally, after this response categories aggregation, separate Rasch analysis is performed and Table 3 shows some summary results of model fitting (Item-trait Interaction and Reliability Indices), obtained applying RSM separately office by office. Chi-squares tests is statistically significant only for the office E (p-value < 0.05) meaning that there are one or more items that are not coherent with the unique latent trait. Disaggregate values of Chi-square (not reported in Table 3 for brevity) show that item I11 (Clearness and simplicity in forms compilation) is not coherent for office E and should be reformulated.

The value of the Person Separation Index (usually employed to check reliability in RM) suggests that reliability is achieved for each office.

The specific analysis of quality and satisfaction for the five offices suggests essentially to improve the calibration of the questionnaire and to reduce the rating scale from a ten-point to a five-point scale.

TABLE 3. - *Summary results for Rating Scale Rasch Model*

	Office	A	B	C	D	E
Item-trait Interaction	**Chi-square test**	18.807	19.334	17.695	15.464	29.774
	DF	12	14	14	12	14
	p-value	0.09	0.153	0.221	0.217	0.008
Reliability Indices	**Separation Index**	0.918	0.919	0.924	0.918	0.915

Results obtained using Rasch analysis are reported in Table 4 where item ranking with five response categories is compared with previous order obtained using the median-HI criterion.

Order of items doesn't change for office A; only the last two items are inverted for offices D and E; more relevant changes in order are present for offices B and C. But if we consider last items for each office (those that need improvement) the two orders highlight approximately the same critical aspects.

TABLE 4. - *Items ranking using median-HI criterion and Rasch analysis*

Office	Item ranking (from better to worst one)						
	Best	←	←		→	→	Worst
A	**I03**	**I02**	**I01**	**I05**	**I06**	**I04**	
	I03	***I02***	***I01***	***I05***	***I06***	***I04***	
	(-1.044,0.212)	*(-0.666,0.192)*	*(-0.505,0.185)*	*(0.153,0.190)*	*(0.800,0.175)*	*(1.261,0.158)*	
B	**I03**	**I02**	**I08**	**I01**	**I07**	**I04**	**I09**
	I03	***I01***	***I02***	***I08***	***I09***	***I07***	***I04***
	(-0.628,0.124)	*(-0.341,0.121)*	*(-0.306,0.121)*	*(-0.240,0.120)*	*(0.420,0.119)*	*(0.535,0.119)*	*(0.560,0.119)*
C	**I10**	**I11**	**I03**	**I08**	**I01**	**I07**	**I04**
	I08	***I01***	***I03***	***I11***	***I10***	***I07***	***I04***
	(-0.318,0.116)	*(-0.298,0.117)*	*(-0.295,0.118)*	*(-0.045,0.115)*	*(0.057,0.114)*	*(0.447,0.118)*	*(0.452,0.118)*
D	**I03**	**I12**	**I02**	**I06**	**I01**	**I13**	
	I03	***I12***	***I02***	***I06***	***I13***	***I01***	
	(-0.817,0.143)	*(-0.190,0.153)*	*(0.037,0.140)*	*(0.208,0.140)*	*(0.282,0.140)*	*(0.480,0.139)*	
E	**I03**	**I02**	**I01**	**I13**	**I12**	**I11**	**I14**
	I03	***I02***	***I01***	***I13***	***I12***	***I14***	***I11***
	(-0.969,0.127)	*(-0.548,0.124)*	*(-0.010,0.117)*	*(0.007,0.117)*	*(0.199,0.119)*	*(0.559,0.137)*	*(0.661,0.117)*

Note: first line: order with median-HI criterion; second line (in italic) order with Rasch (in brackets: ILP estimation and standard error).

The third step is to analyse the global quality of the service and of the users' satisfaction considering the questionnaires of the five offices all together with the aim to compare performances and satisfaction levels. Questionnaires of the five offices, as already pointed out, are formed by common and unique items. To compare users' satisfaction it is necessary to equating the five different forms of the questionnaires, so that they share a unique common scale. Many methods can be used to perform equating (for a comprehensive review see, for example, Kolen and Brennan, 2004). When different forms of a questionnaire share some items in common, the so called *concurrent calibration* method can be used (Kim and Kolen, 2007). This procedure is performed by merging data set of the five offices all together and by estimating –

in a single analysis – item and person parameters using the new data set. The merged data set we consider is formed by ten items (those that are common to almost two offices), as if a new questionnaire was administrated for the different offices and users have to answer to a subset of items that is partially the same for each office. The merged data set is sketched in Table 5, showing the overlapping items between the five original forms of the questionnaires: total number of items is 10 (each with five responses ordinal categories) and total number of respondent is 555.

TABLE 5. - *Merged data set with 10 items (and rearranged order of offices)*

Item Office	I01	I02	I03	I04	I06	I07	I08	I11	I12	I13	n. of valid responses
A											90
B											112
D											112
E											132
C											109

Concurrent calibration resulting by global RSM is then performed, but it gives unsatisfied results: even if the reliability index is excellent (Separation Index=0.992), the Chi-squares test is statistically significant (X^2=195.023, DF=80, p-value<0.05) meaning that uni-dimensionality seems not to hold. Further investigation would be necessary to better understand results regarding chi-square test. Nevertheless users' satisfaction, measured by person parameters obtained with previous concurrent calibration, is compared with results of the next and final data analysis.

The last step of our analysis regards the comparison of offices performances and the computation of a global satisfaction measure for the service using, respectively, CPI and GSI indexes presented in Section 3. In so doing we take into account results obtained with separate RSM (*i.e.* the new appropriate ordinal five-point scale). CPI values, for each office, as well as the PLP averages obtained with concurrent RSM, are reported in Table 6. As it is possible to see, office A is the one with the higher satisfaction level while office C is the one with the lower. Looking at values in Table 6, it is also possible to see that offices order is the same for the two analysis (dissimilarity based index and concurrent RSM).

TABLE 6. - *CPI_h values and PLP averages*

Office	*A*	*E*	*B*	*D*	*C*
CPI value	0.8207	0.6953	0.6675	0.6006	0.5385
PLP Average	2.4132	1.1420	1.0985	0.6356	0.5352
(SE)	(1.5391)	(1.8099)	(1.7105)	(1.6504)	(2.0303)

Note: in brackets: PLP standard error.

Finally, the value of the GSI index (that resumes the overall level of satisfaction) is equal to 0.6645 (that is about the 66% of its maximum).

5. CONCLUDING REMARKS

This paper suggests complementary use of different methods to analyse users' satisfaction about services from PA. Since surveys frequently involve corporations that supply different services to different users, the main difficulty is to choice the methodology to treat data collected in disaggregate samples to perform not only a disaggregate analysis, but also an overall evaluation. The complementary methods reported in this paper can be used to achieve these two goals. Separately analysis can be conducted with *scorecard* ranking methods and with a separate RSM (for questionnaires calibration and scale check). A global analysis is performed through concurrent RSM and dissimilarity indexes.

ACKNOWLEDGEMENT

The authors want to express their thanks to Mario Passon, chief of the Statistical Office of the Chamber of Commerce of Udine, and Elena Marton for her admirable job of collecting data.

REFERENCES

Andrich D., Sheridan B., Luo G. (2004). *RUMM 2020, version 4.0*. RUMM Laboratory Pty Ltd, Perth.

Bertoli-Barsotti L. (1999). Assessing consumer perceptions of service quality by the polytomous Rasch model: existence and uniqueness of the ML estimates. In *Valutazione della qualità e customer satisfaction: il ruolo della statistica*. Vita e Pensiero, Milano, 255-273.

Bond T.G., Fox C.M. (2001). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences.* Lawrence Erlbaum Associates, London.

Capursi V., Porcu M. (2001). La didattica universitaria valutata dagli studenti: un indicatore basato su misure di distanza fra distribuzioni dei giudizi. In Atti del Connvegno Intermedio SIS 2001, *Processi e metodi statistici di valutazione*, (Roma 4-6 giugno 2001), Università di Roma ''Tor Vergata'', Roma.

Cronin J.J. Jr., Taylor S.A. (1992). Measuring Service Quality: A Reexamination and Extension. *Journal of Marketing*, **56**, 55-68.

Fischer G.H., Molenaar I.W. (1995). *Rasch Models. Foundations, Recent Developments, and Applications*. Springer-Verlag, New York.

Giudici P. (2007). Governo dei Rischi: Il ruolo dei modelli statistici. *Istituto Lombardo (Rend. Lett.),* **141**, 361-376.

Kim S., Kolen M.J. (2007). Effects on Scale Linking of Different Definitions of Criterion Functions for the IRT Characteristic Curve Methods. *Journal of Educational and Behavioral Statistics,* **32**(4), 371-397.

Kolen M.J., Brennan R.L. (2004). *Test equating, linking and scaling: Methods and practices* (2nd Ed.). Springer-Verlag, New York.

Kotler P. (1978). *Al servizio del pubblico*. Etas Libri, Milano.

Leti G. (1983). *Statistica descrittiva*. Il Mulino, Bologna.

Pagani L., Zanarotti M.C. (2003). Analisi della qualità di un servizio: un confronto tra scale mediante il modello di Rasch. *Statistica & Applicazioni*, **2**, 35-54.

Pagani L., Zanarotti M.C. (2010). Some uses of Rasch Parameters in Customer Satisfaction Data Analysis. *Quality Technology and Quantitative Management*, **7**, 83-95.

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**, 12-40.

Rampichini C., Grilli L., Petrucci A. (2000). Analisi della qualità della didattica attraverso modelli multilivello, in Civardi M. e Fabbris L. (Eds.), *Valutazione della didattica con sistemi computer-assisted*, Cleup, Padova.

Rasch G. (1960). *Probabilistic models for some intelligence and attainment tests*. Danish Institute for Educational Research, Copenaghen.

Tanese A., Negro G., Gramigna A. (Eds.) (2003). *La customer satisfaction nelle amministrazioni pubbliche. Valutare la qualità percepita dai cittadini.* Rubbettino, Roma.

Zaccomer G.P., Marton E. (2009). L'indagine di Customer Satisfaction camerale: un caso di: un caso di studio per una differente prospettiva di analisi. *Rivista di Economia e Statistica Rivista di Economia e Statistica del Territorio*, **3**, 76-99.