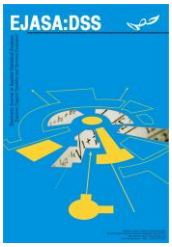

Electronic Journal of Applied Statistical Analysis: Decision Support Systems and Services Evaluation
EJASA:DSS (2010). Vol 1, Issue 1, 26 – 41
e-ISSN 2037-3627, DOI 10.1285/i2037-3627v1n1p26

CUSTOMER SATISFACTION WITHIN A LARGE SCALE RETAIL TRADING GROUP: RESEARCH DESIGN AND CONTROL

Antonino Mario Oliveri*, Franco Vaccina
Dipartimento di Metodi Quantitativi per le Scienze Umane,
Università degli Studi di Palermo, Italy

Received 06 February 2010. Accepted 12 October 2010
Available online 17 December 2010

Abstract*: Even in contexts where Customer Management techniques are not very advanced, the need for relations with customers is as strong as to suggest expensive* ad hoc *researches able to simultaneously measure the satisfaction levels, to detect the eventual inadequacy of selling actions, to investigate the relations of customers with competitors. This is why the* Grande Migliore *Group in Palermo commissioned a research from the University of Palermo. The research was made in year 2008 at 10* Grande Migliore *shops in Palermo and in other Western Sicily towns. The multipurpose nature of the research and the features of the customers population suggested the adoption of an* ad hoc *research design, the features and control of which are described in this paper with particular focus on the impact of some non sampling errors.*

Keywords*: Customer satisfaction; large scale retail trading; research design; non sampling errors; interviewer effect.*

1. Customer Satisfaction (CS) researches and the new frontiers of marketing

Ad hoc researches on CS, carried by means of interviews or self-administered questionnaires, have in recent years been largely implemented within the Large Scale Retail Trading sector, with the purpose of reinforcing relations with customers and augmenting loyalty, with special reference to the technological sectors, that represent the firm core business.
The purpose of knowing clients is actually pursued by means of complex strategies, within which *ad hoc* researches represent just one of many data sources on customers behaviours and motivations.

*Corresponding Author. Email: oliveri@unipa.it

The real aim is that of assembling the information fragments concurring to the customer knowledge, making them available when it is useful to the firm functions; once unthinkable computer science resources are now usable to this purpose.
In the 90's the implementation of ERP (*Enterprise Resource Planning*) systems [11] allowed to manage fundamental firm processes (production, administration, marketing, etc.) as a whole; that was the starting point for new procedures today permitting the best equipped firms to take advantage of the sole knowledge of their customers, about which many data are available and can be correlated.
It is the base of the Customer Relationship Management (CRM) approach [14] [6].
In the absence of suitable and well designed data warehouses, and when personal data of customers and purchase traces over time cannot be related[1], firms need information about customers' opinions. In such conditions, the limited purpose of superseding some traits of the customers' *continuum* is substituted by the more general aim of obtaining a wide-ranging view of customers' satisfaction levels, buying behaviours and motivations.
In section 2 the main features of the CS sample research are presented. In section 3 sample design and sizes are presented, while in sections 4 and 5 the controls on the data gathering process are discussed with reference to some non sampling error sources. In the last section some conclusions are drawn.

2. The survey on CS at the *Grande Migliore* Group in Palermo

The *Migliore* Group works in Palermo since the 30's. In the 90's it opened new department stores and specialized shops in Palermo and in other towns in Western Sicily. Within its shops the group actually proposes a supply ranging from bathroom fittings to toys, gifts, technologic systems, hi-fi products.
With the purpose of measuring CS, the *Grande Migliore* Group charged the University of Palermo with a survey, made from April to July 2008. For each shop belonging to the Group, the research team measured CS levels, detected weaknesses in the supply of services, investigated the contacts of customers with competitors.
In the planning phase, the research team faced many problems related to the demands of the Group and to available resources. They can be synthesized as follows:

1) the multi-purpose nature of the survey;
2) the large number of data gathering points and their territorial distribution;
3) the demand of the Group to obtain results interpretable at the level of single shops;
4) the need, due to limited available budget, to employ only six interviewers and for a limited time interval (twelve successive weeks);
5) the identification of the sampling frame to be used in stratification procedures and for the calculus of the related sampling error.

With reference to point 1), we decided to distinguish the analysis of the service quality from that of the customers' behaviours/buying motivations (with direct reference to contacts with competitors). The need to bound the length of interviews and to simplify the data gathering forms suggested to conduct separate surveys: a *Services* survey and a *Products* one. Each survey,

[1] If such relations can be established, CS researches can be addressed to the analysis of no-buying behaviours.

independent of the other, was made within each of the 10 shops belonging to the Group, using separate interview forms which differed also depending on the kind of shops. After a pilot research, made during December 2007, we finally formulated (and used) 6 different interview forms (*A-B-C Products* Forms, *A-B-C Services* Forms).
The results of the pilot research suggested to modify the wording of some questions and to homogenise the behaviours of interviewers, in order to reduce their impact on the final researches. For instance, interviewers were suggested to reduce or increase the length of interviews in order to approach the average, ranging from less than 4 minutes at the specialised shops to little more than six minutes at department stores.
The fact that only one interviewer obtained, at one of such stores, 75% of total refusals (almost all by she-customers), was pointed out and specific training was prescribed to him.
Refusals were caused by "hurry" and were especially given by adults 36-65 years old. Response rates always exceeded 70% (at some shops 90%) and this was considered a favorable result respect to the effectiveness of the strategy arranged in order to intercept customers.
Points 2-5 above refer to the issue of the sampling design arrangement.
Since no lists of actual customers were available, and effective customers were considered the only eligible statistical units (having already selected a product and keeping it in own hands or putting it in the shopping basket), they could not be intercepted at home (no telephone numbers nor postal addresses were known), nor at the shop entrance (impossible to foresee who, inside the shop, would have bought at least a product); nor the exit would fit (after having paid, clients do not look so permissive about interviews). We consequently decided to contact customers inside the shops. Each interviewer had to move following a typical path through all departments, stopping when his/her individual timetable prescribed an interview. At that moment, he/she had to select the nearest effective customer on his/her right, in both round trip paths. In the unfortunate case of a refusal, the next nearest customer should have been selected.
Interviews were made at all shops and time slot, all the days of the week. Interviewers were rotated, in order not to leave any shop undone in any day or time slot.

Table 1. The interviewers rotation plan

Weeks of the year	Days of the week	Interviewer 1	Interviewer 2	Interviewer 3	Interviewer 4	Interviewer 5	Interviewer 6
Week 16	Monday	1	6	5	4	3	2
	Tuesday	2	1	6	5	4	3
	Wednesday	3	2	1	6	5	4
	Thursday	4	3	2	1	6	5
	Friday	5	4	3	2	1	6
	Saturday	6	5	4	3	2	1
Week 17	Monday	2	1	6	5	4	3
	Tuesday	3	2	1	6	5	4
	Wednesday	4	3	2	1	6	5
	Thursday	5	4	3	2	1	6
	Friday	6	5	4	3	2	1
	Saturday	1	6	5	4	3	2

Weeks of the year	Days of the week	Interviewer 1	Interviewer 2	Interviewer 3	Interviewer 4	Interviewer 5	Interviewer 6
Week 18	Monday	3	2	1	6	5	4
	Tuesday	4	3	2	1	6	5
	Wednesday	5	4	3	2	1	6
	Thursday	6	5	4	3	2	1
	Friday	1	6	5	4	3	2
	Saturday	2	1	6	5	4	3
Week 19	Monday	4	3	2	1	6	5
	Tuesday	5	4	3	2	1	6
	Wednesday	6	5	4	3	2	1
	Thursday	1	6	5	4	3	2
	Friday	2	1	6	5	4	3
	Saturday	3	2	1	6	5	4
Week 20	Monday	5	4	3	2	1	6
	Tuesday	6	5	4	3	2	1
	Wednesday	1	6	5	4	3	2
	Thursday	2	1	6	5	4	3
	Friday	3	2	1	6	5	4
	Saturday	4	3	2	1	6	5
Week 21	Monday	6	5	4	3	2	1
	Tuesday	1	6	5	4	3	2
	Wednesday	2	1	6	5	4	3
	Thursday	3	2	1	6	5	4
	Friday	4	3	2	1	6	5
	Saturday	5	4	3	2	1	6
Week 22	Monday	1	6	5	4	3	2
	Tuesday	2	1	6	5	4	3
	Wednesday	3	2	1	6	5	4
	Thursday	4	3	2	1	6	5
	Friday	5	4	3	2	1	6
	Saturday	6	5	4	3	2	1
Week 23	Monday	2	1	6	5	4	3
	Tuesday	3	2	1	6	5	4
	Wednesday	4	3	2	1	6	5
	Thursday	5	4	3	2	1	6

Weeks of the year	Days of the week	Interviewer 1	Interviewer 2	Interviewer 3	Interviewer 4	Interviewer 5	Interviewer 6
	Friday	6	5	4	3	2	1
	Saturday	1	6	5	4	3	2
	Monday	3	2	1	6	5	4
	Tuesday	4	3	2	1	6	5
	Wednesday	5	4	3	2	1	6
Week 24	Thursday	6	5	4	3	2	1
	Friday	1	6	5	4	3	2
	Saturday	2	1	6	5	4	3
	Monday	4	3	2	1	6	5
	Tuesday	5	4	3	2	1	6
	Wednesday	6	5	4	3	2	1
Week 25	Thursday	1	6	5	4	3	2
	Friday	2	1	6	5	4	3
	Saturday	3	2	1	6	5	4
	Monday	5	4	3	2	1	6
	Tuesday	6	5	4	3	2	1
	Wednesday	1	6	5	4	3	2
Week 26	Thursday	2	1	6	5	4	3
	Friday	3	2	1	6	5	4
	Saturday	4	3	2	1	6	5
	Monday	6	5	4	3	2	1
	Tuesday	1	6	5	4	3	2
	Wednesday	2	1	6	5	4	3
Week 27	Thursday	3	2	1	6	5	4
	Friday	4	3	2	1	6	5
	Saturday	5	4	3	2	1	6

Products research	*Services* research

1=Shops A,G; 2=Shop D; 3=Shops B,H; 4=Shop E; 5=Shops F,I,L; 6=Shop C.

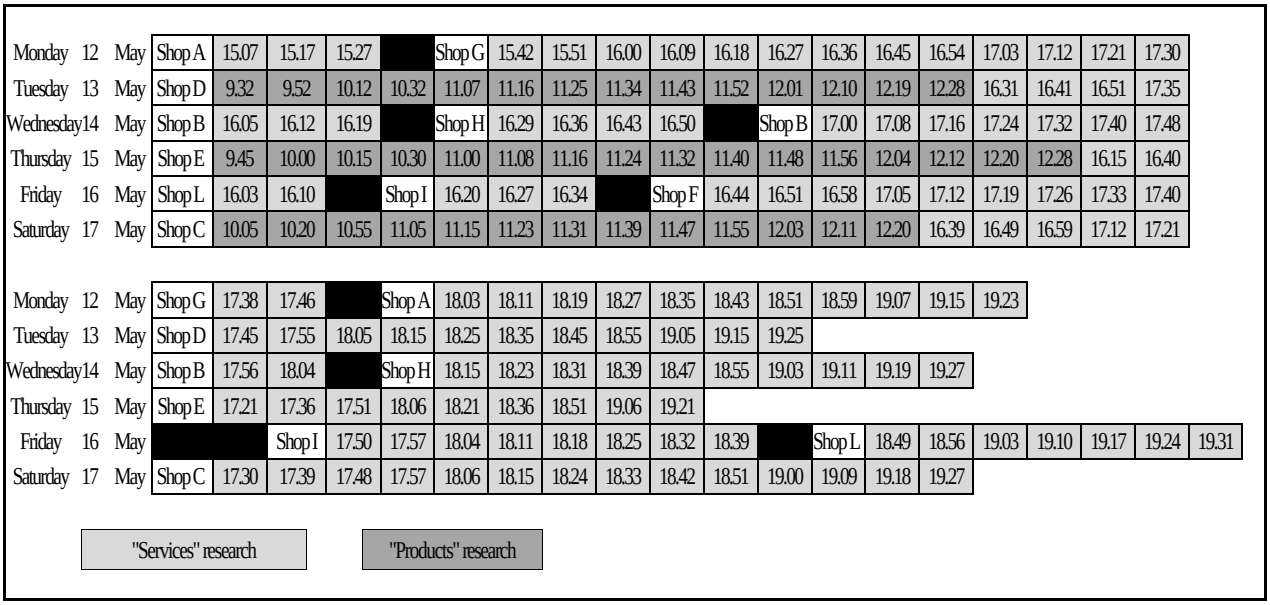

Figure 1. Interviewer 5 individual timetable, 12-18 may 2008

For each survey (*Products* and *Services*) and shop the planned number of daily interviews is the result of proportional stratification by week and time slot, with reference to the traffic volumes of receipts (*proxy* of the number of visits) in year 2007.
Layers were determined so as to balance weekly and daily periods of *high* and *low* traffic (if greater or lower than the average); the reason resides in the opportunity of not overestimating the foreseeable unfavourable effect of crowding on satisfaction, nor the favourable effect of less crowded periods.

3. Sample design and size

In order to obtain results interpretable at each shop level, the research team planned to extract 20 independent samples: one for the *Products* and one for the *Services* survey for each of the 10 shops.
Respect to the multi-purpose nature of our researches, a procedure as that suggested by Cochran [3], consisting in the selection of the most relevant variables respect to which calculating the optimal sample size, could have been followed but the final obstacle was the absence of requested information on which "the most important" variables and what their distributions were.
It was also considered that, even if a stratified sample should perform better than a simple random sample in terms of the efficiency of estimates (favourable design effect), the expected gain is often not considerable yet.
The 20 simple sample sizes were finally determined by means of equation (1):

$$n = \frac{\frac{z_{\alpha/2}^2 PQ}{d^2}}{1 + \frac{1}{N}\left(\frac{z_{\alpha/2}^2 PQ}{d^2} - 1\right)} \qquad (1)$$

d=0,05 being the maximum deviation of the estimate *p* from the parameter P [3, p. 75], *α=0,05*. For each shop *N* was determined on the base of 2007 data.
The above reported considerations suggested us to adopt prudential solutions:

a) with reference to the multi-purpose nature of the researches, it was decided to determine sample sizes suitable to all aims, that is by fixing *P=0,5* in equation (1);
b) given the fact that stratified sampling should require smaller samples than simple random sampling, sizes were determined referring to the latter.

Equation (1) was applicable since, for the sake of simplicity, polytomous variables originally ranging (in 4 categories Likert-like scales) from "completely satisfied" to "not at all satisfied", were transformed into dichotomous ("satisfied" - "unsatisfied"). This choice is justified by the fact that respondents were actual customers to be considered "satisfied" only when "completely" such. All other categories were consequently collapsed into the "unsatisfied" one.
As Table 2 shows, at the end of the data gathering phase not all expected sample sizes were fulfilled, due to the difficulty of reaching the planned ceilings within less crowded shops, especially in low crowded time slots. Nor the enlargement of contacts to all present customers permitted to work the problem out.

Table 2. Planned and obtained interviews, by shops and research

Shops	***"Products"* research**		***"Services"* research**	
	Planned	***Ceiling attainment (Yes/No/Almost)***	***Planned***	***Ceiling attainment (Yes/No/Almost)***
A	384	Y	384	Y
B	383	Y	383	A (329)
C	383	Y	383	A (367)
D	384	Y	384	Y
E	376	Y	376	A (348)
F	376	N (284)	376	N (256)
G	380	A (337)	380	Y
H	350	N (183)	350	N (160)
I	355	N (170)	355	N (134)
L	374	N (254)	374	N (248)
Total	3745		3745	

In such specialised next-door shops interviews were made in the same days and subsequent times, yet belonging to the same time slot. Interviewers had consequently no time enough to wait for the arrival of new customers.
The populations of specialised shops are less heterogeneous and quite smaller than others yet. For such shops (in Table 3 and Figure 2: shops F, H, I, L), the sampling and response rates are quite favourable.

Table 3. Sampling rates (time interval: april-july 2008, estimation made on the base of 2007 data)

Shops	Ratio $(n_i/N_i)*100$	
	"Products" research	**"Services" research**
A	0,05	0,05
B	0,18	0,15
C	0,33	0,32
D	0,11	0,11
E	2,08	1,92
F	1,64	1,48
G	0,89	1,01
H	4,72	4,13
I	3,60	2,84
L	1,78	1,74
Total	0,20	0,19

Note: in grey the shops for which a smaller number of interviews was observed, in comparison with that planned

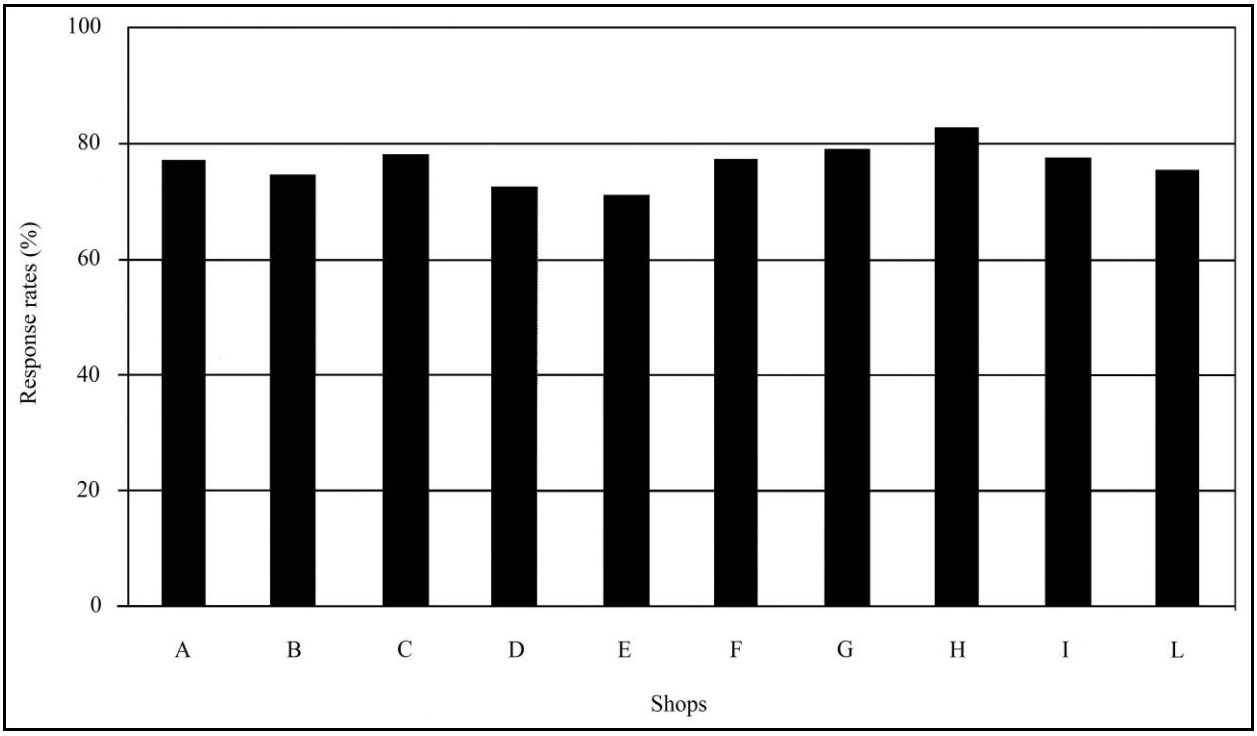

Figure 2. Response rates, by shop

Budget and time saving demands prevented the research team from increasing the shifts numbers for interviewers; this solution would have modified, on the other hand, the sampling plan in terms of a lack of control on the error. Even if the performed census data gathering within some space-time cells could increase total error, it reduced that referred to specific subsamples.
The limited increase in sampling error for specialized shops was predictably repaired by reduction in non-sampling errors amounts.

In Table 4 expected maximum sampling errors, calculated on the base of the actual sample sizes (Table 2) and the simple random sampling, are reported for each shop and research.
The original sampling design of our research, approximating the stratified without-replacement design, should yet produce smaller sampling errors (a consequence of prudential solutions we adopted).
At the Grande Migliore Group level the overall sampling error reduced to an amount smaller than 2%.

Table 4. Estimates of maximum expected absolute sampling errors, by research and shops (95% confidence level, *P*=*Q*=0,5)

Research	Shops										All shops
	A	B	C	D	E	F	G	H	I	L	
Products	±0,048	±0,047	±0,050	±0,046	±0,049	±0,058	±0,053	±0,071	±0,074	±0,061	±0,017
Services	±0,047	±0,054	±0,051	±0,047	±0,052	±0,061	±0,050	±0,076	±0,083	±0,062	±0,018

4. Controls for the quality of the data gathering process. CAPI-PAPI comparisons

As reported in Figure 2, response rates (expressed as percentages) were always around or more than 75%; the analysis of responses, complete and coherent for almost 99%, showed high reliability of data. Figure 3 shows that the mean number of unsuccessful contacts preceding a valid interview was always less than 1,5 (much less in most cases). Such results suggest that the selected strategy for contacting clients proved to be effective.

Figure 3. Mean number of unsuccessful contacts, by shop

Particular attention was paid to some relevant non-sampling error sources. Of the five sources of non-sampling error identified by Biemer and Lyberg [1]: specification, frame, non response, measurement, processing, the last three were specifically controlled.
Interviews were in fact made with the aid of paper or electronic forms according with a *quasi-experimental* design aiming at assessing the differential effects of face-to-face paper (PAPI) and computer assisted (CAPI) interviewing in terms of efficacy, efficiency and the production of non-sampling errors [12]. In extremely synthetic terms, a number of randomly selected CAPI units were assigned to one group, and a corresponding number of PAPI units were assigned to the control group by matching. Cases and controls were reciprocally as homogeneous as possible with reference to some control variables potentially affecting responses: age, sex, conditions under which the interview was made: presence/absence of children, time slot (few/many customers in the shop), shop category and department.
With reference to the whole unit non response error (measured by the non response rate), CAPI was expected better than PAPI, as well as with reference to the setting errors (the computer should increase the respondent feeling of higher privacy level). Preliminary results do not show that the two tools really provided different performances (that is: no more accurate data were produced by the CAPI tool), respect to both the interview dynamics (measured for instance through the respondents nuisance or intolerance) and sensitive questions (such as: *do customers think products are warranted by being sold into a particular shop?*) [13].

5. Non-sampling errors: estimates of the interviewer effect

Respect to PAPI, CAPI permits more control on interviewers, due to the automatic recording of information about each interview (as instance: opening and closing time). As in PAPI, also in CAPI the interviewer has a great impact on the survey, especially if inexpert in managing the electronic device. At the end, the interviewer's role seems the real critical element also in CAPI researches. In order to minimize the impact of such a source of potential systematic errors, within our research the basic training course for interviewers included interview software and computer handling.
When speaking of the "interviewer effect", we refer to increases in the variance of estimates (overdispersion), due to correlation among respondents interviewed by the same individual.
The interviewer effect can be controlled by means of proper data gathering designs. The most common technique consists in the *compenetration of assignments* [10], that is the constitution of equivalent subsamples of respondents assigned to different interviewers. Compenetration can be obtained by randomization and this is what we pursued through the complete rotation of the interviewers respect to shops, days of the week, time slots.
The following considerations will be made with reference to the *A-Products* form (Shops A, C, D). As reported in Table 5, interviewers completed around the same number of interviews.
With respect to the distribution of interviews by interviewers-dependent *respondents groups* (explanatory variable) and *day of the week* (response variable), some deviations from what had been planned can be observed; no relationship can yet be detected between the two variables by means of the asymmetric Goodman's and Kruskal's λ index [7], usually employed for nominal variables: $\lambda = 0{,}05$. With respect to the response variables *time slots* and *shops*, deviations are

less evident and, once again, no relationship with *respondents groups* appears ($\lambda \cong 0{,}00$; $\lambda = 0{,}026$).

Table 5. Distribution of interviews by interviewers, *A-Products* form

Interviewer code	Number of interviews	%
1	226	17,795
2	213	16,772
3	204	16,063
4	216	17,008
5	196	15,433
6	215	16,929
Total	1270	100,000

The interviewer effect is usually taken into account when:
1) respondents' assessments or attitudes are requested;
2) the respondent is requested to make a relevant effort;
3) ambiguous questions are formulated.
The first case is very common in our research. Assessments and attitudes were measured on nominal or ordinal scales; in such cases the overdispersion related to the interviewer effect cannot be modelled by means of the usual ANOVA approach [9] (the condition of normality in the distribution of errors cannot, for instance, be verified [2] [15]).
In general, the interviewer effect consists of a multiplier of the estimator variance and can be defined as:

$$Deff_{\text{int}} = \left[1 + (m-1)\rho_{\text{int}}\right] \qquad (2)$$

ρ_{int} being the so-called "intra-interviewer correlation coefficient" that can be estimated on the base of relations between *inter* and *intra* interviewer variances within the ANOVA environment [8] [9], while m is the supposed constant workload assigned to interviewers (when it is not, m can be substituted by $\overline{m}$, the mean number of interviews *per* interviewer).
When the ANOVA assumptions cannot be met, the class of logistic regression models represents a valid option. In such cases, a random effect component is usually added to the fixed one yet, in order to model overdispersion. The probability of observing a certain answer category, supposing that it can vary among the groups of respondents intercepted by different interviewers, is modelled according to certain distributional laws [4, p. 206]:

$$logit\left(\vartheta_i\right) = \underline{x}' \underline{\beta} + \delta_i \qquad (3)$$

As usual, in Equation (3) ϑ_i, $\underline{x}$, $\underline{\beta}$ respectively indicate response probabilities, explanatory variables and regression parameters, while the δ_i term represents the random effect, $i = 1,2,...,n$.

The random effect is often modelled in terms of the beta distribution, the *a* and *b* shape parameters of which are commonly considered in order to estimate the intra-interviewer correlation coefficient.
The beta function can be expressed as follows:

$$B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}\,dx \tag{4}$$

with mean and variance of *X* equal to $\frac{a}{a+b}$ and $\frac{ab}{(a+b)^2(a+b+1)}$ respectively.
As well known, in the presence of overdispersion, within a binomial logistic model the variability in the response probabilities can be expressed as [4, pp. 192-193]:

$$Var(\vartheta_i) = \phi p_i q_i \tag{5}$$

ϑ_i being response probabilities $(i = 1,2,...,n)$, $p_i = E(\vartheta_i)$, $q_i = (1 - p_i)$. $\phi \geq 0$ is a scale parameter. ϑ_i are the values of an unobservable random variable; p_i can yet be estimated through the observed proportions $\frac{y_i}{n_i}$.
If the variance of the response probabilities is modelled according with a beta distribution, it can be shown that [4, p. 204]:

$$Var(\vartheta_i) = p_i q_i / [a_i + b_i + 1] = p_i q_i \tau_i \tag{6}$$

and, in the frequent case that a_i and b_i are assumed constant,

$$\tau = \phi = \frac{1}{a+b+1} \tag{7}$$

Being directly related to random variation in the response probabilities, the ϕ scale parameter is commonly defined *the overdispersion parameter* and consequently usually interpreted in terms similar to ρ_{int} [2] [15].
The interviewer effect can so be estimated according with the (2), where the ρ_{int} parameter is substituted by ϕ.
In Table 6 the calculus of the interviewer effect is reported, with reference to different kinds of questions. Fits were obtained using the Modified Newton Raphson algorithm [5]. ϕ values correspond to the scale parameter estimates.

Table 6. Estimates of the interviewer effect, *A-Products* form

	Questions	Valid responses	Parameters	Estimates	Standard errors	p-values	$Deff_{int}$*
Assessments, attitudes	Are you in general satisfied with the prices within this shop?	1181	Constant	-0,0072	0,4580	0,9875	13,445
			Interviewers	0,0041	0,1175	0,9724	
			Scale parameter	0,0595	0,0362		
	Are you in general satisfied with the assortment within this shop?	1088	Constant	1,4539	0,3572	< 0.001	5,556
			Interviewers	-0,0081	0,0901	0,9283	
			Scale parameter	0,0218	0,0152		
	Are you in general satisfied with the products update within this shop?	1146	Constant	1,2250	0,4468	0,0061	10,615
			Interviewers	-0,1428	0,1114	0,1996	
			Scale parameter	0,0491	0,0308		
	Are you satisfied with the prices within this shop with only reference to hi-fi products?	1078	Constant	0,4566	0,1801	0,0112	1,721
			Interviewers	-0,0843	0,0463	0,0687	
			Scale parameter	0,0040	0,0055		
	Are you satisfied with the assortment within this shop with only reference to hi-fi products?	747	Constant	1,1091	0,4378	0,0113	9,341
			Interviewers	-0,0041	0,1105	0,9701	
			Scale parameter	0,0439	0,0280		
	Are you satisfied with the products update within this shop with only reference to hi-fi products?	747	Constant	1,2624	0,4567	0,0057	10,041
			Interviewers	-0,1184	0,1135	0,2969	
			Scale parameter	0,0506	0,0320		
	Are you satisfied with the prices within this shop with only reference to this department?	682	Constant	0,8779	0,4655	0,0593	7,088
			Interviewers	-0,0865	0,1178	0,4628	
			Scale parameter	0,0493	0,0330		
	Are you satisfied with the assortment within this shop with only reference to this department?	1226	Constant	1,0351	0,3832	0,0069	4,779
			Interviewers	-0,0761	0,0966	0,4310	
			Scale parameter	0,0306	0,0223		
	Are you satisfied with the products update within this	1250	Constant	0,9589	0,4652	0,0393	7,050
			Interviewers	-0,1099	0,1179	0,3513	

	Questions	Valid responses	Parameters	Estimates	Standard errors	p-values	$Deff_{int}$*
	shop with only reference to this department?		Scale parameter	0,0537	0,0357		
Assessments, attitudes	Do you think that this shop is a warranty respect to the quality of products?	1226	Constant	0,5545	0,4242	0,1912	10,943
			Interviewers	-0,0506	0,1082	0,6400	
			Scale parameter	0,0489	0,0305		
Sensitive questions	When you need some goods, do you usually control prices in many shops?	1250	Constant	-0,0973	0,4219	0,8176	10,890
			Interviewers	-0,0664	0,1098	0,5457	
			Scale parameter	0,0477	0,0301		
Final questions	During the interview, did the respondent show nuisance or intolerance?	1264	Constant	1,9402	0,4906	< 0.001	8,150
			Interviewers	0,0106	0,1228	0,9313	
			Scale parameter	0,0341	0,0224		
	Did answers provided by the respondent sound reliable?	1266	Constant	2,3197	0,3029	< 0.001	2,050
			Interviewers	-0,0542	0,0759	0,4750	
			Scale parameter	0,0050	0,0061		

**Multiplier of the estimator "proportion" variance*

In the case of the variable reported in Table 7, the algorithm did not converge, so a fixed effect logistic model was fitted. The close to one odds ratio value suggests small yet significant interviewer effect.
On the contrary, and not coincidentally, Table 6 shows that modeling overdispersion through the beta component makes interviewer fixed effects on the response variables not significant.

Table 7. Estimates for the fixed-effect model, *A-Products* form

	Question	Parameters	Coefficient	Std.Error	p-value	Odds Ratio
Sensitive question	Do you buy in other shops (of a kind similar to this one)?	Constant	0,0805	0,1290	0,5325	1,0838
		Interviewers	0,1355	0,0342	< 0.001	1,1451

As easily verifiable, even very small ϕ coefficients produce relevant multiplier effects on the variance of estimates, greater and greater as the sample size (and consequently the workload for interviewers) increases. This is not surprising yet, being consistent with its non sampling error nature. The large amount of the interviewer effect is, on the contrary, to be interpreted since we must consider that the detected within-subsamples relevant homogeneity should be completely attributed to interviewers only, in the absence of possible effects produced by concurrent factors; one of such factors being, for instance, the eventual (and not easily distinguishable!) failure of the attempt to constitute equivalent respondents groups by means of randomization (this is why interviewers were rotated and strict rules regarding units selection were prescribed). As known, randomization represents in fact a warranty against the introduction of systematic distortion in the constitution of groups, but in no case can warrant that, after constitution, groups are effectively equivalent. The estimated multiplier effect attributed to interviewers so represents the upper limit of an interval, the lower limit being zero.

6. Some conclusions

The research on CS made at the *Grande Migliore* shops was the occasion to think about the impact of non-sampling error sources on sampling researches. The larger impact of such errors with respect to sampling error has often been detected, as much greater as a) the number of interviews increases and b) the number of interviewers is small. This unfortunately happens in the great majority of researches, where condition (a) is commonly (but unjustifiably) considered a desirable aim, and condition (b) is a frequent constraint. The latter was also the case of our research, for which a specialized agency was entrusted by the firm with the task of selecting and controlling interviewers. On one hand, the fact that some already selected people were replaced with others before beginning the survey, did not permit proper training; on the other hand the budget constraints imposed by the firm management did not permit to increase the interviewers number. Such elements contributed to determine the high interviewer's effect respect to many investigated variables.

The sole interventions at the research group's disposal were: a) the preservation of the random selection of statistical units, necessary to keep at least the sampling error under control, and b) to quantify the interviewer effect by means of the compenetration of assignments through the complete rotation of interviewers.

For such aims, an original sampling design was arranged which permitted to conclude 20 contemporaneous researches at 10 different shops, obtaining results interpretable at the single shop level.

The mode effect of different data gathering tools was controlled too through a (quasi) experimental design, and no substantial improvement was found about CAPI over PAPI performances.

We are aware of the fact that when strong constraints are imposed over empirical researches and these are more dependent of the market than of scholars demands, non sampling errors cannot easily be avoided, but just assessed and quantified in order to protect the purchaser firm managers from the inaccurate employment of results in eventual support of tricky decision making processes. Anyway, more and more conscious effort has to be produced so as to put under control as much amount of non sampling error as it is possible, in order to obtain more realistic estimations of the total error [16].

Acknowledgements

This paper is supported by Università degli Studi di Palermo ex-60% funds within the research project: *The Service Charter: contexts and methods for its building, check and control in Sicily*, directed by A.M. Oliveri, and by the proceeds of an agreement between the *Grande Migliore* Group and the Dipartimento di Metodi Quantitativi per le Scienze Umane at the Università degli Studi di Palermo. Although the work is common responsibility of both authors, sections 1 and 6 can be attributed to Franco Vaccina, sections 2, 3, 4, 5 to Antonino Mario Oliveri.

References

[1]. Biemer, P., Lyberg, L. (2003). *Introduction to Survey Quality*. New York: Wiley.
[2]. Biemer, P., Stokes, S.L. (1991). Approaches to the modelling of measurement error. In *Measurement errors in surveys*, eds. Biemer P.P., Groves R.M., Lyberg L.E., Mathiowetz N.A., Sudman S., New York: Wiley.
[3]. Cochran, W.G. (1977). *Sampling Techniques*, 3rd edition. New York: Wiley.
[4]. Collett, D. (1994). *Modelling binary data*. London: Chapman and Hall.
[5]. Dennis, J.E., Schnabel, R.B. (1983). *Numerical methods for unconstrained optimization and nonlinear equations*. New Jersey: Prentice-Hall Inc.
[6]. Duse, M. (2009). *Il CRM strategico. Come migliorare la competitività aziendale fidelizzando e centralizzando il cliente*. Milano: Franco Angeli.
[7]. Goodman, L.A., Kruskal, W.H. (1954). Measures of association for cross-classifications. *JASA*, 49, 732-764.
[8]. Hansen, M.H., Hurwitz, W.N., Bershad, M.A. (1961). Measurement errors in censuses and surveys. *Bulletin of the ISI*, 38 (2), 359-374.
[9]. Kish, L. (1962). Studies of interviewer variance for attitudinal variables. *JASA*, 57, 92-115.
[10]. Mahalanobis, P.C. (1946). Recent experiments in statistical sampling in the Indian Statistical Institute. *JRSS*, 109, 325-378.
[11]. Marchi, M., Mancini, D. (eds.) (2003). *Gestione informatica dei dati aziendali*. Milano: Franco Angeli.
[12]. Oliveri, A.M., Parroco, A.M. (2008). The Analysis of Large-scale Retail Trade Groups Customer Satisfaction: A Comparison Between CAPI and PAPI Data Collection Methods. In *Proceedings of the 7th Intern. Conference on Social Science Methodology*, Napoli, 1-5 September, 2008.
[13]. Oliveri, A.M, Parroco, A.M (2009). CAPI Versus PAPI and The Production of Some Non-Sampling Errors. In *Proceedings of CLADAG 2009 Conference*, Catania, 9-11 September, 2009, Padova: Cleup, pp. 589-592.
[14]. Peppers, D., Rogers, M., Dorf, B. (2000). *Marketing one to one. Manuale operativo del marketing di relazione*. Milano: Il Sole 24Ore Pirola.
[15]. Searle, S.R., Casella, G., McCulloch, C.E. (1992). *Variance components*. New York: Wiley.
[16]. Vaccina, F. (1980). Evitabilità e inevitabilità dell'errore. In *Manuale del colloquio e dell'intervista*, eds. Trentini G., Milano: ISEDI, cap. 9.