

Sampling Issues and Management Solutions in Internet-Based Market Researches

M. V. Ciasullo, G. Festa

Abstract. *In the broad context of internet-based market research, this paper focuses on research carried out on net users. An analysis of the advantages, but also disadvantages, related to online market research shows that a crucial problem is the 'vagueness' of the representativeness of sample surveys, even affecting the credibility of the online survey. The authors have developed a theoretical / practical framework that integrates contributions from marketing, statistics and information technology, following a logical-methodological path oriented towards reducing uncertainty in online surveys. The model utilizes helpful considerations from inferential statistics and proposes solutions for managing internet samples as virtual communities. The conceptual paper aims to contribute to an advancement in the field of online market research, shifting the focus on the validity / reliability of the investigation from the mere software or statistical tool to a wider analytical internet marketing strategy.*

Keywords: e-research, analytical internet marketing, online sampling.

1. Introduction

The *customer based* approach [Valdani and Busacca, 2000], fundamental for the success and survival of any organization, has become a prevalent trait of corporate behavior nowadays. In this respect needs analyses are essential if we are to grasp a customer's way of thinking and the means by which she/he chooses or otherwise, to buy. At the same time, an ever more globalized economy, characterized by the capacity to rapidly transfer information, delineates not only more attentive and complex customer profiles, but also new styles of consumption based on the capacity of the goods/products to reflect the culture, taste and style of each individual.

The ongoing changes associated to growing markets and their related innovative dynamics in terms of purchasing, impose on businesses constant commitment in the analysis and interpretation of the above phenomena. It is clear therefore, that in the context of studies on marketing management, particular attention has been addressed to the analytical aspect of marketing, for the purpose of analyzing the structure of the market and trends in demand in order to identify the most significant factors that impact on consumer behavior.

Market research, generally intended as systematic and objective analyses for the purpose of gathering relevant information in order to elaborate and put in place interaction strategies with clients and other members of the corporate value system, observe markets and delineate scenarios. Therefore, it is evident how such strategies constitute basic drivers of corporate action, where competitiveness is based on the capacity to conquer and foster ready, exigent and innovative demand.

Within such a complex context, the value of high tech prevalently *internet-based* and capable of contributing significantly to social research and specifically market research finds its place. The Internet is a suitable 'natural' tool for gathering targeted, in-depth and appropriate information relative to demand [Kiang et al., 2000], both in a market research context and prior to that, through pertinent operations of *marketing intelligence* [Kotler, 2001]. In such a perspective, an important element of the web is underlined i.e. its capacity to render communicational processes substantially independent from the limits of time and space. Evidently, this attribute is particularly valuable with respect to research online in that it enables the overcoming of a series of barriers that might exist between researchers and the parties taking part in the survey.

Furthermore, as evidenced in the literature [Cantone et al., 2006; Vescovi, 2007], new information and communication, in particular based on open, universal standards, take on a fundamental role in improving the management of the intra and inter-business relational system and of the underlying processes of value creation, facilitating at the same time, the phases of approach to international culture and markets. Consequently, the successful outcome of market research depends to a great extent on the constant updating combined inevitably with evolving *internet-based* marketing information and, in the context of our study, the strategies of analytical internet marketing in the broadest sense.

2 Internet-based market research categories, taxonomies and critical issues

The combination between marketing and statistics is traditionally at the basis of the skills system regulating market research [Bassi, 2008]. Marketing in this context, uses inferential statistics prevalently in the compiling of the sample for the survey stage in order to extend sample findings to the population. Over time, these two skills have branched into a third discipline i.e. information technology which by virtue of the benefits involved in terms of time and costs, has contributed remarkably, to the two skills mentioned above, to the extent that it can be said that the market research system is governed by the sum of these three disciplines (marketing, statistics and information technology).

The Internet on the other hand, has not only engendered a new series of information technology tools for researchers, but by virtue of its success in 'social' terms has ended by offering the market research field, a new strand of investigation. Thus, it is now possible to distinguish between market research carried out *by means of* the Internet and market research directly in contact with

web users *on* the Internet: the two categories making up the extended family of *internet-based marketing research*.

In the context of the present study, our focus will not be on the former category (*by means of*) primarily in the sense that we consider the Internet exclusively a tool and in the second place because it has already been analyzed by the Authors in a previous work [Festa, 2001b]. In this study we intend to concentrate on the second category (*on the net*), by identifying the main critical issues of reference (i.e. the representativeness of the sample) and proposing where possible, a theoretical analysis that takes into account the integrated contribution of marketing, statistics and IT.

We consider that devising appropriate methodology relative to the issue of *internet-based* market research has to start from a traditional market research approach. This is necessary in order to verify whether the relative 'toolbox' of techniques (i) is suitable for internet use, (ii) applicable if adapted, using IT mode and (iii) applicable in an innovative key, exploiting the enormous potential of the web. Two factors in epistemological terms underpin this point of view [Festa, 2001a]: the 'technological' characteristics of the web platform and the 'social' characteristics of the web users. For the sake of brevity they are analyzed by reviewing the advantages and disadvantages linked to market research on line, without presuming to examine the issue in any depth.

As concerns some of the main advantages, prevalently in terms of efficiency, economy and timesaving (where current IT hardware and software enable elaborations to be made in real time) and reduced costs (not in an absolute sense but if compared to the normal costs of traditional marketing research) the Internet becomes in a certain sense, consecrated as a more efficient channel for data collecting [Weible and Wallace, 1998]. In any event, even more significant qualitative advantages of the Internet result. The potential of the web however, should not be considered merely in terms of stored data but also and above all, with respect to the manner of collecting, which offers the interviewee a more pleasant experience (thanks to multimedia applications), greater independence in terms of time and place of data detection. Another factor is the convenience and accuracy of detection for both parties together with more independence from the context (personal and/or situational for categories of persons such as the disabled, the ill, the secluded or in areas such as war zones or zones where epidemics are rife etc.), thus breaking down psychological and socio-cultural barriers (pivoting on the sensation or mere virtual presence of the other parties) [Schmitz, 2004; Wright, 2005].

The overall sense of convenience and comfort, furthermore, generally produces more satisfying results especially as concerns qualitative research which seems quite a propitious and interesting investigating scenario for *internet-based* market research [Andreani and Conchon, 2001b]. In this respect, new competences are required of researchers e.g. the skills required of a moderator of a group online. Qualitative research aspires to attaining in-depth understanding of consumer behavior by analyzing the aspirations and sensations associated for example, to a category, brand or image [Mariampolski, 2001; Marbach, 2010]. Besides statistic representation

(fundamental in quantitative research), qualitative techniques have the advantage of applicability even in cases where information is difficult to detect, complex to analyze or of problematic interpretation [Andreani, 1998; Tissier-Desbordes, 1998; Andreani and Conchon, 2001a] as they offer the interviewees greater freedom of expression [Prandelli and Verona, 2006]. Thanks to the Internet however, the interviewee can take over 'command' of the interview, thus preventing the interviewer from 'manipulating' the interview in any way. If this were not the case and should the interviewee become aware, a click would be sufficient to log out or even remain online fooling the interviewer with trick responses. Consequently, technological, methodological and above all socio-relational skills in line with the Internet environment will be required of the online qualitative research designer and relative interviewer. For this reason too in our opinion, the most interesting market research on the Internet is qualitative research, precisely because it enables the extrapolating of information which in other contexts an interviewee would be hard put to confess (despite the fact that non-verbal language might easily give him away). It goes without saying however, that many disadvantages can be linked to market research online, fully compensated for nonetheless by the above mentioned advantages. For instance, the extent to which a user is familiar with the 'language' of the keyboard or mouse (i.e. monitoring a chat line) or the loss of information on the context (i.e. a questionnaire online to which one responds either from one's office or from home, being 'bound' naturally enough to the rules of behavior regulating different environments) and not least, to the impossibility of grasping the indications of non verbal language (or even those of language: different way of expression using electronic writing compared to using traditional methods or even voice mode, register or tone of voice etc.). In any event, as emerges from many studies [Howard et al., 2001; Andrews et al., 2003; McDonald e Stewart, 2003; Di Fraia, 2004; Wright, 2005] the main defect of internet based research is linked to the 'vagueness' of the representativeness of the survey sample which can even jeopardize the credibility of the survey itself. As the literature confirms this weakness, the main aim of our conceptual paper is to develop a theoretical / practical framework that can better delineate the boundaries of the online sample representativeness issue, proposing specific guidelines for governing the same deriving from complex analytical internet marketing strategies. In this respect, the implications of the particular insights we offer could be the basis for further research on the subject.

3. Elements characterizing online sample surveys

Analyzing the problems relative to online sample surveys, with a vision based on 'gaps', not all the reference population are necessarily regular users of the Internet (*netizen*); not all can necessarily be *identified / contacted* by the researcher for the survey; even among those contacted, not all of them would necessarily decide to respond (*respondent*); and among these not all would necessarily want to respond honestly (*authentic*) or with trustworthiness (*reliable*), the reasons being the spatial-temporal gap/distance in terms of the survey and sense of 'irresponsibility' socially circulating online. The number of web surfers for instance using free services (e-mail, video sharing, freeware,

etc.) who when requested by the provider to register for commercial purposes, give unreliable or even false personal data in order to overcome the obstacle of registration and to use the online services is huge. Such obstacles tend to create an effective *funnel* which is illustrated in Figure 1.

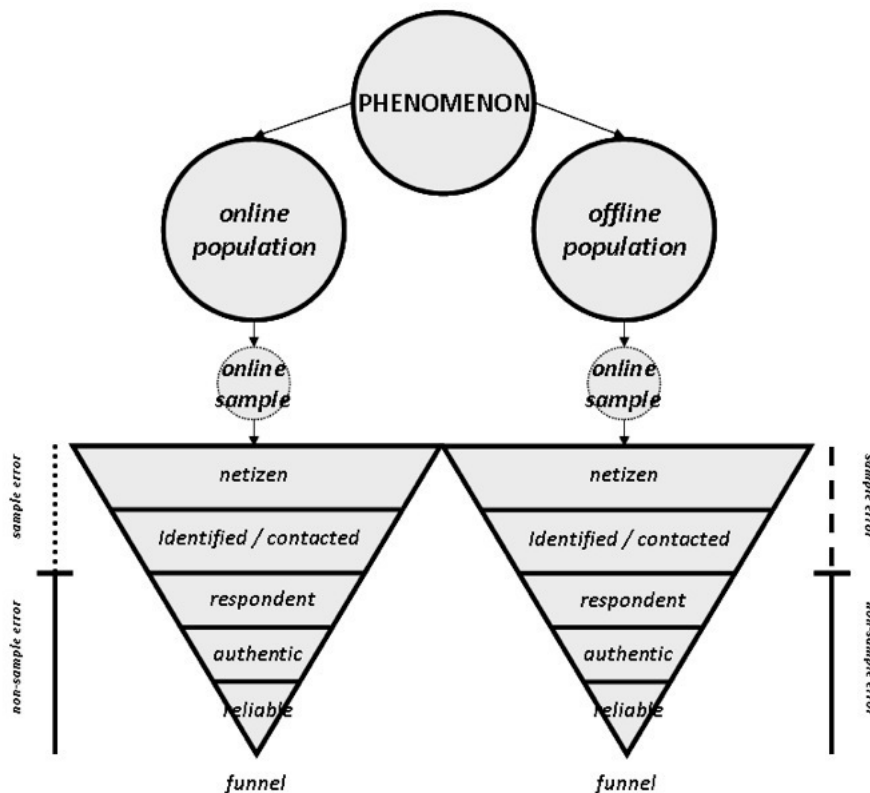


Figure 1
Common sample and non-sample errors in online surveys

The figure clearly shows that internet based market research involving both the online or offline populations, presents identical problems in terms of non sample errors (i.e. errors on the part of interviewers, interviewees, researchers etc.), where *respondent*, *authentic* and *reliable* have to be governed but only 'from a bottom up perspective' (i.e. the group of individuals that can be identified and contacted on the Internet). Previously, the question of representativeness was posed underpinning the rationale of our study or in other words, the potential (almost certain) gap between the offline population and the sample online. This contributes to rendering the sample error even more complex, as shown by the bold line in Figure 1.

Sample errors strictly speaking are in the main, identical for both samples (this observation constitutes one of the presuppositions for our methodological

approach), as it derives from the application of the central limit theorem, according to which should the number involved in the sample increase, representativeness also increases (or, more precisely, with the increase in the number involved in the sample, distribution of the sample tends towards standard thus enabling reliable predictions once errors and representativeness have been defined) [Natale, 2004]. Nonetheless, it is clear that a sample online is the 'legacy' of a population online, while in the case of a sample online used for representing a population offline, such legacy results inevitably, spurious.

In our view, this appears to be a major issue of online research as not only is it impossible to classify the devices used in *internet-based market research*, as online tools are in constant evolution (technological and social) above all their combination, but from the point of view of marketing, such an effort could in effect, result neither decisive. On the contrary, representation of the sampling problem and relative solution could be devised following a conceptual scheme which starts from the logical-methodological framework of marketing research and proceeds by means of a synoptic vision of the tools and – above all – techniques available for use.

In this context, following a strictly classical approach to marketing research taking into account the flexibility needed to meet various professional scenarios, it is fundamental to start from a solid conceptual basis. As a result, a suitable methodological approach to *internet-based* market research set in the framework of the traditional process of market research, would entail eight stages [Barile and Metallo, 2002]:

1. defining the marketing issue;
2. identifying the aim of the research;
3. formulating the aims of the survey;
4. designing intervention (in terms of quality, times and costs);
5. economic evaluation of scheduled actions;
6. data collecting (primary and/or secondary, from internal and/or external sources);
7. data processing and analyzing (with suggestions for application);
8. final reporting (even for accreditation).

Obviously such an approach cannot be considered mandatory but the advantages in schematic terms are evident. Proceeding with the structuring of the approach, the most delicate phase concerns 'designing intervention'. This phase can be divided into eight sub-phases [*ibid.*]:

1. formulating operative research objectives;
2. planning the survey in practical terms;
3. designing the sample;
4. selecting detecting tools;
5. managing human resources engaged in the survey;
6. scheduling time scales for activities;

7. planning results tabulating and codifying;
8. cost estimating.

This kind of division is typical of the logical-methodological process of market research. The first sub category for instance, (*formulating operative research objectives*) and the last (*cost estimating*) of the 'designing intervention' phase are linked to the previous phase (*formulating the aims of the survey*) and to the subsequent phase (*economic evaluation of scheduled actions*).

Diverse conceptual 'zones' characterized by identical problems both in the *earth-based* and *internet-based* research (i.e. analysis of problematic issues in marketing) emerge clearly from the above. Other 'zones' on the contrary, regard the tools context (i.e. questionnaires online as opposed to paper documents), obviously more advantageous for market research online. Finally a 'zone' characterized by specific issues regards internet-based market research, the most relevant of which is undoubtedly representativeness (as mentioned above and confirmed from a methodological perspective).

Observation (cf. the *funnel*) and methodology (cf. the *literature review*) show that the main issue relative to *internet-based* research consists in the limited representativeness of the survey. To overcome such limit, in our view, it is considered fundamental the support of marketing to inferential statistics, the opposite of what usually occurs in traditional market research where the contribution of statistics is functional to the analysis and eventually to the solution of the marketing problem [Molteni and Troilo, 2003].

Math and statistics which constitute the disciplines contributing most to the study of sample representativeness, cannot but derive from the GIGO rule ('garbage in, garbage out') [Seglin, 1994] and in a certain sense, show no interest for the quality of the starting data. On the contrary, the researcher well aware of the potential of the Internet has to examine this issue in depth in an attempt to overcome contextual limits. Currently, the core issue is the representativeness of the online sample with respect to phenomena (also) offline. While it is evident that survey of a particular phenomenon online could pose serious problems of representativeness if carried out offline, or in other words, involving interviewees who are not normally users of the internet services under analysis.

Paradoxically, therefore, we show that online research, although deeply affected by the 'funnel', cannot be considered unreliable beforehand. It would be epistemologically wrong for example to investigate a user by means of traditional techniques even only in terms of a *concept test*, relative to the experience of browsing in a multimedia virtual store which could even be 3D. Some proposals are developed in this perspective in the concluding part of the paper where the methodology of collecting 'authentic' and 'reliable' data from the *internet sample* conceived as virtual community [Hagel and Armstrong, 1997], is discussed relative to the problem of marketing in question.

4. Potential solutions for the problem of online representativeness

In this section the main characteristics of an adequately represented sample is discussed and accounted for statistically, with, in coherence with our study, the main focus on managerial aspects. Generally, the representativeness of the sample (n) with respect to the population (N) is delivered, together with the correct calculation of the sample number, by means of probabilistic sampling, in the first place by mere random sampling [Molteni and Troilo, 2003]. Only random sampling (i.e. the odds known and the same for each member of the sample) can guarantee the absence of selection errors (common on the contrary in non probabilistic sampling, i.e. by convenience, opinion, quota). Eliminating selection errors from random sampling together with an adequate sample numbering should ensure the absence of distortion.

In non-probabilistic surveys, furthermore, researchers are not unaware of potential distortions but on the contrary, use them to their own advantage. They consider that a deliberate error of selection enables a more adequate contextual representativeness although this obviously cannot be known beforehand. From random extraction a more representative non-probabilistic sample could emerge, because random with respect to the same (identical) non-probabilistic sample itself. The paradox lies in the fact that the random sample finds its strength not in the result but in the process: furthermore, even a random sample (precisely because it is random) can produce 'exceptional' outcomes.

The sample as mentioned previously, is based on the central limit theorem according to which the increase in number of the sample renders the sample distribution closer to standard thus allowing calculation of the required number of the sample to support a given margin of error (E_{s_x}) and a given trust/confidence interval (z). The formula set out below illustrates how (see Table 1) the sample number is calculated (a population in excess of 100.000 elements is defined infinite by statistic convention, a phenomenon is binomial if represented in only two modes, continuous in other scenarios) [Barile and Metallo, 2002].

	Finite population	Infinite population
Continuous phenomenon	$n = N z^2 s^2 / [(N - 1) E^2 s_x + z^2 s^2]$	$n = (z s / E_{s_x})^2$
Binomial phenomenon	$n = N z^2 p (1 - p) / [(N - 1) E^2 s_x + z^2 p (1 - p)]$	$n = p (1 - p) (z / E_{s_x})^2$

Table 1
Calculating sample number (n) compared to number of population (N)

Estimated accuracy therefore, depends not so much on the fraction of the sample ($f = n / N$) but rather on the number of the sample by virtue of the strength of the central limit theorem. Such considerations, crucial in sampling theory, enable us to state that the issue of representativeness on the Internet to a certain extent, is inspired more by the practical aspect of the survey rather than

by the scientific aspect of the methodology. In other words, the first tier of the funnel can be ignored, i.e. the non-coinciding of the population under observation with that of the Internet users as long as a sufficiently numerous online sample (updated statistically in terms of mortality rate) can be made up.

In practice, the most relevant issue concerning online representativeness refers to the number of elements of the sample concurring to form an adequately calculated sample number. The fact that such elements of the offline population also apply to online users represents an additional as opposed to penalizing characteristic. On the contrary, recruiting exclusively active online users for a sample of offline population could to some extent be assimilated to non probabilistic sampling [Andreani and Conchon, 2001a], deriving from opinion or convenience (or even by quota: e.g. a survey measuring customer satisfaction of a bank having both traditional branches and internet based services).

In non probabilistic sampling, the underpinning concept of standard error in the calculating of sample number has absolutely no value. Qualitative market research tends prevalently to make use of non probabilistic sampling, which also for this reason seem to be the most practicable online market research survey, driven by definition, not by descriptive but merely exploratory aims. Given that sample representativeness should be governed by the number of internet users surveyed as opposed to the extent of internet users compared to the population, it should not be necessary therefore to extract an online sample from an online population corresponding in part or as a whole to the offline population seeing as the positive confirmation of the characteristics of the population in the survey would be sufficient provided it were adequate in number. The procedure would therefore not be that of extraction (from top to bottom) but of abstraction (from bottom to top).

It is fundamental, therefore, to mitigate data distortions linked to the survey by means of credentials in terms of transparency: characteristics of the population, type of sampling, number of sample, percentage of errors acceptable and estimated reliability. All the other information considered of impact for outcomes should be added to the statistic indications (in terms of research integrations): considerations on the training of the interviewers, on the formulating of the questions, on the classifying of responses and so on, addressing attention to putting in place correctly systems for the eliminating or mitigating of non sampling online error, with particular reference to the 'funnel' phases relative to *respondent*, *authentic* and *reliable* categories.

Such credentials represent not only the 'litmus test' for interpreting the findings of the survey but also and above all, the key for evaluating information obtained. Consequently, the credentials themselves become information for the decision making process. The natural statistic distortion of the sampling, due to the effective distance or gap between population and sample cannot be questioned: in the same way that it stands for *earth-based* research, it certainly stands for *internet-based* (the socio-economic nature is constantly changing and internet users are people). Consequently, the issue of representativeness of the internet population with respect to offline events constitutes the main focus of this study. Our logical-methodological reflections outlined above would seem

to suggest a solution (taking into account characteristics and context involved). By means of a global strategy of analytical internet marketing, aimed at constructing, monitoring and governing the internet samples as if they were virtual communities, would render it possible to discuss useful devices for guaranteeing beforehand the 'clarity' of incoming data, resolving in this way, the further limits imposed (represented by the *funnel*) utilizing our proposed solutions outlined below.

5. Guidelines for managing an internet sample

On the basis of the theories discussed in our study, some fundamental methodological (*strategic*) and applicative (*operative*) characteristics could be delineated not limited merely to appropriate planning but also to the efficient managing of the internet sample for online research. In this sense the idea of a sort of virtual community appears particularly appealing the construction of which should be based on strategic internet marketing as opposed to the perspective of market research in the strictest sense and underpinned (even when analytical) on the binomial 'visibility' and 'attraction' [Metallo and Festa, 2003].

In other words, paradoxically, the focus should not be on the issue of the validity and reliability of online market research as such, but rather that of the number of population online, from which to extract the sample base and subsequently, the sample for the survey. The latter if adequate numerically speaking might be sufficient in terms of representativeness at least from an empirical point of view: from the methodological perspective on the contrary, the above approach might not be practicable, not being in substance possible to proceed on the Internet to a random sampling which is certainly representative of an offline population [Schmitz, 2004]. In this direction, the *funnel* requires constant adjusting and mitigating. In other words, insistence should be not only on the statistic planning/design of the sample but also and above all, on the correct management of the sample, whereby a series of devices is activated to constantly authenticate interviewee data for the purpose of guaranteeing identity and authenticity in data detection.

However, on the other hand, such barriers could at the same time provoke the opposite effect whereby the interviewee, irritated by being subjected to constant authentication processes, becomes less spontaneous and natural in responding and even tires of being interviewed altogether. To date in any event, this seems to be the only path to follow to ascertain the reliability of the survey online. Various authentication steps are necessary and can be illustrated as follows:

- *registration*: the user has to register with the virtual community (the researcher consequently, right from the start of the survey, has to imagine the sample online of any research online as a virtual community, thus deriving a greater sense of 'belonging' on the part of the interviewee in terms of the positive outcome of the research), compiling a registration form, on the basis of which the researcher can compile a series of preliminary data in order to classify the user in conformity with the principal criteria of segmentation of the community. Such strata are neither static nor dynamic but merely 'virtual' (in the same way that in a database, the *queries*

can be considered virtual tables), exploiting the capacity of the IT system underpinning the virtual community to extract, and filter the most useful segments in real time (OLAP applications). On registration, the system assigns the user a *username* and *password*, giving them authenticated, authorized and responsible access to the system;

- *accreditation*: during the process of data detection, users might be requested not only to authenticate their identity (using the account) but could also be subject to other authorizations through for example a *captcha* (on the access page of the survey) and/or a *password* (in the e-mail of 'invitation to take part in the survey' sent to the user). *Captcha* is the acronym for 'completely automated public Turing test to tell computers and humans apart' to confirm the use of the internet service on the part of an individual and not a machine. The user is required to digit a set number of characters displayed on the page into an empty box for verification purposes. The characters seem distorted but an individual can read them easily (a further aim is to raise the level of interviewee attention). Thus the system guarantees against the user (or the IT system underpinning the virtual community) being unwittingly infected by *malware*, which infecting the platform could create conditions for unreliable responses;
- *verification*: to make sure it is effectively the user involved in the survey, it is vital to oblige the same to provide adequate responses to specific questions that only the user can give. Such questions should be presented in random mode in order to 'oblige' the particular user to interact with the system. The process of user authentication represents a core element of the IT security infrastructure and knowledge based techniques are currently the most frequently used to authenticate user identity [Ellison et al., 2000; Dhamija and Perrig, 2000]. Specifically, by means of *knowledge based authentication* (KBA) the user in order to be recognized as such, has to respond accurately to a (standard) question, the response to which has been memorized in the IT system on a previous occasion. This helps to create greater empathy with the user, who should then see herself / himself not as a merely passive 'number' in the battery of interviewees but rather as an individual being asked to interact dynamically. This technique is widely used to allow users to regain access to internet services in the event the password is forgotten. Should the response to the standard query be correct, the automatic system sends an e-mail to the address previously provided by the user reminding her / him of the username and password (or engendering new ones). Some queries in effect could be conceived as 'control' questions in order to obtain the maximum reliability possible in the interviewee's responses [De Luca, 2006].

As mentioned previously, the system of constant authentication could slow down or even irritate the user. However, to date no more incisive or efficient system seems to exist. Other mechanisms of access are in place, based not on what the user 'knows' as analyzed up to now, but rather on what the user 'owns' (e.g. a smart card, a token, etc.) and/or on what the user 'is' (biometric scanning) [Teti and Festa, 2009]. Even in such cases however, a user could easily gain access

to the system for the first time with her / his own credentials and leave someone else to take part in the survey at a later date. The process of constant authentication (registration - authentication - verification) refers consequently to an ongoing verification process relative to user behavior in terms of seriousness (eliminating or mitigating problematic issues concerning the related stages of *respondent*, *authentic* and *reliable* categories) of the 'funnel'. As in all monitoring activities, from an operative perspective, performance is slower; at the same time, however, improved quality in terms of the survey are guaranteed. It could happen that some of the selected users abandon the survey or even the internet sample managed as a virtual community: at this point a compromise between sample quality and sample quantity should be sought. For this reason, qualitative research seems more adequate from an internet environment perspective being less sensitive to sampling problems and sample number and being aimed primarily at exploratory and not descriptive aims.

6. Research findings, implications and conclusions

The evolving of IT and in particular, the introduction, diffusion and development of the Internet have had, continue to have and no doubt will have in the future an extraordinary impact on routine economic and social processes. These innovations have naturally impacted on enterprise information systems above all, provoking radical changes in procedural design, working and development (organizational processes) and logics (information flows) called upon to carry out specific business activities.

One of the disciplines involved to a great extent in such change is that of market research which uses the Internet as an additional tool together with its traditional 'toolbox' (research 'using' the Internet) or even as an additional/substitute/innovative target of individuals to survey (research 'on' the Internet). The most problematic issues from a research perspective derive effectively speaking from research carried out 'on' the Internet where – besides the inevitable need to contextualize adequately new IT tools – potential distance (social, cultural, economic, etc.) has to be taken into account between the population offline and that online, with inevitable repercussions on the appreciation of the representativeness of the online sample compared to the population offline.

In this study, clearly oriented towards research of a conceptual nature, we have defined a logical representation of the main problematic issues of sampling in online research and developed a theoretical-practical framework ('funnel') which takes into account on the whole, the sample and non sample errors which could emerge in methodological terms in the representativeness link (on the one hand) between the population online and sample online and (on the other) between population offline and sample online. By virtue of this scheme, with the support of marketing, statistics and IT resources, we have suggested some empirical solutions for the managing of online representativeness, over and beyond 'substantial' statistics in the case of sampling error and over and beyond 'applied' marketing in the case of non sampling error.

These solutions in any event are useful only if pursued in the prospect of the internet sample of interviewees conceived as a virtual community, considered by

the researcher (and above all by themselves) as members of a specific active online group even if only for the duration of the survey. Thus it is fundamental to devise an analytical internet marketing strategy in order to adapt and maintain the sample / community, resorting to the changing combination of visibility and attraction.

The research implications of the present study seem to concern above all the potential evolution of the *humus* at the foundation of the market research discipline or in other words, the traditional combination of marketing, statistics and (by now) IT, identifying in such mix of skills, grounds for further research both in methodological and applied terms. By virtue of the pervasiveness of the Internet, in depth studies on market research will need in the future to deal with the social evolution of the Net which is already scheduled to meet (more so in the future) the need / opportunity of offering solutions (for the present) and/or far-sighted solutions (for the future), in the same way that the reflective spirit of this study identifies indications for application above all in terms of operative implications and therefore of interest for anyone called upon to carry out research online.

Market research online it has to be noted is still at an extremely early stage. This is generally true for all internet-based environments as concerns values, principles and above all working rules (e.g. netiquette, privacy, social networks, etc.). This characteristic, together with the vertiginous speed of development of internet tools and applications and their potential of use which compress the duration of the *internet year* exponentially, renders already obsolete or at least, risky, any empirical solution (one of the most likely limits of our conceptual paper). For this reason, it is fundamental to devise, contextualize and develop a strict logical-methodological framework such as that underpinning our study. Our approach, starting with the analysis of the 'traditional' methodology of market research, proposes, in the prospect of the Internet and subsequently with regard to the prospect of sampling on the Internet, substantially 'common sense' management solutions typical of business management scenarios. In other words, our study goes beyond the mere analysis of the usefulness of a particular statistic or IT tool, but measures considerations and examines applications observed as respondent to a global analytical internet marketing strategy.

References

- Andreani J.C., Conchon F., Gli studi qualitativi in Internet, *Micro & Macro Marketing*, 1, 2001a, 65-74.
- Andreani J.C., Conchon F., *Les études qualitatives en marketing*, Paris: ESCP-EAP, Les Cahiers de Recherche, 2001b, 01-150.
- Andreani J.C., L'interview qualitative en marketing, *Revue Française du Marketing*, 1988, 3-4.
- Andrews D., Nonnecke B., Preece J., Electronic survey methodology: A case study in reaching hard-to-involve Internet users, *International Journal of Human-Computer Interaction*, 16, 2, 2003, 185-210.
- Barile S., Metallo G., *Le ricerche di mercato. Aspetti metodologici e applicativi*, Giappichelli, Turin, 2002.
- Bassi F., *Analisi di mercato. Strumenti statistici per le decisioni di marketing*, Carocci, Rome, 2008.
- Cantone L., Calvosa P., Testa P., Tecnologie ad alta intensità connettiva, relazioni di marketing e processi di creazione di valore per i clienti nei mercati BtB, *Mercati & Competitività*, 2006, 3, 1.38.
- De Luca A., *Le ricerche di mercato. Guida pratica e metodologica*, Angeli, Milan, 2006.
- Dhamija R., Perrig A., D'èj`a Vu: A User Study Using Images for Authentication, *Proceedings of the 9th USENIX Security Symposium* Denver, Colorado, USA, August 14-17, 2000, 1-15.
- Di Fraia G., Validità e attendibilità delle ricerche on line, in Di Fraia G. (ed.), *e-Research. Internet per la ricerca sociale e di mercato*, Laterza, Rome-Bari, 2004.
- Ellison C., Hall C., Milbert R., Schneier B., Protecting secret keys with personal entropy, *Future Generation Computer Systems*, 16, 2000, pp. 311-318.
- Festa G., Il ruolo dei nuovi strumenti informatici nel sistema informativo di marketing, *Esperienze d'impresa*, 2, 2001b, pp. 67-114.
- Festa G., *Digital Business English - Glossario ragionato di linguistica d'impresa per la new economy*, Edisud, Salerno, 2001a.
- Hagel J., Armstrong A., *Net gain*, Harvard Business School Press, Boston, Massachusetts, USA, 1997.
- Howard P., Rainie L., Jones S., Days and nights on the Internet: The impact of a diffusing technology, *American Behavioral Scientist*, 45, 3, 2001, 383-404.

- Kiang M.Y., Raghu T.S., Huei-Min Shang K., Marketing on the Internet — who can benefit from an online marketing approach?, *Decision Support Systems*, 27, 2000, 383-393.
- Kotler P., *Marketing Management*, Pearson Education Italia, Milan, 2001.
- Marbach G., *Ricerche per il Marketing*, Utet, Turin, 2010.
- Mariampolski H., *Qualitative Market Research*, Sage, London, UK, 2001.
- Martone D., Furlan R., *Online Market Research Tecniche e metodologia delle ricerche di mercato tramite Internet*, Angeli, Milan, 2007.
- McDonald H., Stewart A., A comparison of online and postal data collection methods, *Marketing Intelligence & Planning*, 21, 2, 2003, 85-95.
- Metallo G., Festa G., La progettazione dei portali web nell'ottica del customer service, in Barile S., Metallo G. (eds.) *Soluzioni problematiche d'impresa. Riflessioni e modalità risolutive*, Edizioni Culturali Internazionali, Rome, 2003.
- Molteni L., Troilo G., *Ricerche di marketing*, McGraw-Hill, Milan, 2003.
- Natale P., *Il sondaggio*, Laterza, Bari, 2004.
- Prandelli E., Verona G., *Marketing in rete, oltre Internet verso il nuovo Marketing*, McGraw Hill, Milan, 2006.
- Schmitz N., Rappresentatività e campionamenti nelle survey on line, in Di Fraia G. (ed.), *e-Research. Internet per la ricerca sociale e di mercato*, Laterza, Rome-Bari, 2004.
- Seglin J.L., *Marketing (la guida al)*, McGraw-Hill, Milan, 1994.
- Teti A., Festa G., *Sistemi informativi per la sanità*, Apogeo, Milan, 2009.
- Tissier-Desbordes E., Les études qualitatives dans un monde post-moderne, *Revue Française du Marketing*, 1998, 3-4.
- Valdani E., Busacca B., Customer Based View: dai principi alle azioni, Conference Proceedings "Le tendenze del Marketing in Europa", Università Ca' Foscari Venezia, 2000, 1-24
- Vescovi T., *Il Marketing e la rete. La gestione integrata del web nel business. Comunicazione, e-commerce, sales management, business to business*, Il Sole 24Ore, Milan, 2007.
- Weible R., Wallace J., Cyber research: the impact of the Internet on data collection, *Market Research*, 1998, 10, 3, 19-31.
- Wright K.B., Researching Internet-Based Populations: Advantages and Disadvantages of Online Survey Research, Online Questionnaire Authoring Software Packages, and Web Survey Services, *Journal of Computer-Mediated Communication*, 10, 3, 2005.

Biographies

Maria Vincenza Ciasullo is an Associate Professor of Management at the University of Salerno (Italy), where she teaches business administration and corporate governance. Ph.D. in business economics (c/o the University of Naples "Federico II"). Main topics of her studies are management and governance systems in private and public organizations, focusing on corporate social sustainability and business ethics. She successfully combines theoretical and empirical research. She is author of several national and international publications. email: mciasullo@unisa.it

Giuseppe Festa is an Assistant Professor of Management at the University of Salerno (Italy). He has taught several courses, such as economics and management of health organizations and economics and management of information technology. He has published several articles in national and international refereed journals. Internet marketing is one of his most important research topics. email: gfesta@unisa.it