Limits and virtues of a web survey on political participation and voting intentions. Reflections on a mixed-method search path

Faggiano, Maria Paola

Department of Communication and Social Research (CoRiS), Sapienza University, Italy, Rome

Abstract

The Internet offers new opportunities for the empirical research, especially if we consider that nowadays most citizens are made up of web surfers: on the one hand, we are seeing the transfer of some traditional methodologies on Internet, on the other hand we are witnessing the development of new innovative data collection and analysis tools. The study was conducted through a classical survey tool (the questionnaire), using it as part of a web survey. Secondly, we chose Facebook as an instrument which is particularly suitable for the investigated topic (political participation and voting intentions), because the election campaign for the 2018 Italian general election took place, for all parties and candidate leaders, mainly on this Social Network. Two surveys were carried out, the first one in September 2017 and the second one in February 2018, reaching about 850 and 1,400 cases, with similar percentages over the whole block of variables and with stable connections among them. The aim is to highlight the advantages and disadvantages of a Web survey on the topic of political participation, showing particular attention to strategic choices and decisions that impact positively on the data quality, according to a mixed-method approach.

Keywords: web survey; political participation; voting intentions; Social Network; mixed method approach.

1. Old and new sociological survey tools in the Digital Age: A focus on the Web survey

The rapid spread of digital technologies in each dimension of everyday life has inevitably produced changes in practices, styles, relationships and social interactions. The interest of human sciences researchers in the effects produced by the digitalization process is growing; at the same time the Internet offers new opportunities for empirical research (Cipriani, Cipolla, Losacco, a c. di, 2013; Boccia Artieri, a c. di, 2015): on the one hand, we are witnessing the transfer on the Web (and the adaptation process) of some traditional methodologies (primarily the questionnaire), on the other hand we can observe the development of new innovative data collection and analysis tools. The Internet represents a great chance on the political participation side and for many other topics of sociological interest, considering that nowadays the majority of citizens are made up of Internet users, who are so easily reachable for the purpose of compiling a questionnaire (in Italy today 67% of Italians surf the Web – Report of Demopolis, 2017). However, there is still a generational gap, in fact, only young people use the Internet in almost all cases; even if the "digital divide" is gradually being reduced, through the entry into the mass consumption of mobile devices such as smartphones and tablets, also for the elderly population. Instead, it is more and more complex, for the purpose of conducting surveys, to draw up reliable population lists from which to extract random samples, or to acquire updated and complete telephone registers in order to make telephone or face to face interviews. Moreover, we cannot overlook a detail which applies both to online and offline surveys: in many cases excessive circulation of investigations, opinion polls, market research, etc. have made the voter-citizen elusive and uncooperative, always less inclined to be interviewed. Sometimes voters are very suspicious or unavailable to any initiative that requires on his part to answer to a set of questions (sometimes the respondents show that they do not know how to distinguish the request for collaboration to sociological surveys from that one coming from direct marketing activities). There are in fact many online questionnaires: 1. prepared in the launch phase of a new product/service; 2. designed to record the customers satisfaction about a service or to collect the students approval in a school/university; 3. addressed to citizens to know points of view and needs with respect to a collective service; 4. prepared by political parties to mobilize and get to know voters, etc. Focusing specifically on the Internet, although a self-selected sample poses known problems in terms of its statistical representativeness in comparison with its population, the principle of maximum freedom granted to the respondent seems to produce positive effects in terms of fidelity and quality of the collected data and also of the success of a research initiative: a subject is free to choose to participate or not in an investigation, "released" from the presence or voice of the interviewer at the time of compilation, free to choose the moment compiling into a larger time unit. However, despite a careful selection of the platform and the specific virtual space

that hosts a web survey, although there is a continuous monitoring of the phase of data collection by a research team, the participatory outcome of a study condicted on the web is in hands of a rather low number of respondents, considering that the average yield of these surveys is normally less than 1% of the reached contacts. A limited number of completed questionnaires is recorded in many studies in which the online questionnaire was used (Faggiano, 2007; Peruzzi, 2017; Vaccari *et al.*, 2013). Therefore, a sample - more or less consistent and close to the characteristics of the population - will be consist those people who will choose to adhere to the research activity during the period of the data collection. Their participation is based on their available time, their sensitivity and interest, their own skills about the treated topics. Let's dwell on a peculiarity of the online questionnaire, connected with the typical interactivity of virtual environments. It is interesting to note that a respondent can actively participate in the research experience, not limited to the mere filling out his specific questionnaire. Let's imagine an online questionnaire published on a Facebook page: in this case, the respondents can interact via the web with the posts containing the link of the questionnaire, for example by adding a positive or negative reaction towards the initiative of research and/or the sponsoring institution. They can comment, start a debate with the research team and/or other respondents, share or discredit other people's arguments; write privately to the institution that initiated the investigation, contribute to the circulation of the research initiative through online sharing modes.

2. Doing research to evaluate the survey approach: a voters-based study

In making an evaluation on the chosen survey strategy, argued and supported by data, the study we present was conducted through a classic surveying tool such as the questionnaire, using it as part of a Web survey. More precisely, a post containing the link to the questionnaire was published on the Facebook institutional page of the Department of Communication and Social Research of the Sapienza of Rome. The post was previously sponsored in order to reach social profiles as heterogeneous as possible. These profiles were based on social extraction charateristics, professional activity, level of education and hobbies and interests. The post was also shared on other online platforms and channels. The choice of Facebook is obviously not accidental: it is the most used platform in Italy (and not only), with 33% of the registered population, followed by WhatsApp and Messenger (services also widely used by the research team for the purpose of disseminating the questionnaire, in addition to Telegram, Instagram, LinkedIn, Twitter, mailing lists). Moreover, Facebook is particularly suitable for the treated topic (political participation and voting intentions), because the election campaign for the 2018 Italian general election took place, for all parties and candidate leaders, mainly on this Social Network. Compared to the theme of research, Italy has a large tradition of empirical studies, generally carried out with standardized methodologies of questioning of voters (think of the activity of the Italian

Society of Electoral Studies or the Italian National Election studies). On the eve of the 2018 political elections, the dimensions in analysis were: values, sense of legality, idea of social justice, trust in the institutions, social resentment, social problems perceived as urgent, political orientation and electoral behaviour over time, traditional forms of political and social participation, forms of online political participation and hybrid styles. Two fairly close surveys were carried out, one in September 2017 and the other one in February 2018, reaching about 850 cases in the first case and about 1,400 units in the second one, with similar percentages over the whole block of variables and with stable connections among them. The aim is to highlight the advantages and disadvantages of a web survey on the participation topic, showing particular attention to those strategic choices and decisions that have positive effects on the data quality, in a mixed method approach. Although the reached samples cannot be considered representative from the statistical point of view (these are self-selected samples), we want to underline the particular attention used towards the data quality (specially for the phase of design of the data collection tool, the pretest and the data analysis of the first survey, which has been essential to calibrate the questionnaire with a view of its second use and of the improvement of the whole data collection strategy). We have the numbers to analytically describe and deepen opinions, attitudes, values and social practices related to the most significant Italian electoral targets (*left* area, *right* area, Five Star Movement, Non-voting area, area of indecision), in addition to the typical socio-demographic and economic variables. Furthermore, the offline use of the questionnaire (approximately one fifth of the interviews for each survey) has been an effective solution (Kott, Chang, 2010) for reaching a marginal voter who still exists in Italian society (often elderly and with a low level of education): the non-user of the Web (today it is estimated that a quarter of the Italian population does not use the Internet at all).

3. Measures for improving data quality

In a synthetic way, it is useful to spend a few words about all the measures that the research team has developed in order to obtain reliable data and to widen to the maximum the respondents' catchment, not only from the strictly quantitative point of view (the numerical consistency of the sample), but also with reference to its heterogeneity regarding the strategic variables related to the political participation theme (as a problem of statistical coverage). In fact, the Web data collection mode is definable as "device agnostic", that is the questionnaire that we used in the survey is suitable to be compiled either from PC, tablet or mobile device (for example, the scales of attitude originally prepared with scores from 0 to 10, were brought back to a 0-5 range, in order to optimally display each possible mode of response, even from smartphones). Through a meticulous pre-test on a heterogeneous sample (by age and level of education), conducted both online and offline (with a questionnaire in paper format), we made improvements to the wording and the

formulation of the precoded answers. Furthermore, we worked on the order and the number of questions, as well as on the closure of some questions, etc. We have already specify the reasons below the choice of the most used social platform, Facebook, and of the overall strategy of sharing the questionnaire on other channels (Social Networks and mailing lists) in the selected temporal unit (one month for the first and for the second data collection; in the latter case it was the period of the election campaign in a technical sense). With reference to the aspect of sharing, by following the writing of a long list of themes and key words, we tried to identify some Facebook groups connected with social activities and diversified interests, having the electoral targets as a point of reference. It was not easy to penetrate into these groups, especially for increasingly restrictive rules on privacy and spamming. So, an element that we had held in high regard for the success of the survey research, did not help matters much. The surveys conducted in specific groups to which you belong and within which you declare the research activity in an explicit way are very different (case studies, surveys on circumscribed themes, etc.). Many undergraduates in social sciences take advantage of their belonging to youth groups (university, music, etc.) to collect data on lifestyles of their peers; these experiences' success depends both on their role (symmetrical), and on the composition of the target audience (homogeneity compared to a hobby or to age, etc.). Getting the attention of large and heterogeneous populations is more complicated, as it is inevitable that some targets are more receptive than others. The sponsorship of the post, spread throughout the entire research timeframe with a medium economic investment, has led to contain the initial distortion of the sample, that was introduced by a sharing mode of the post based on an university circle and related to the local context of the Lazio Region. This attempt to curb the problem has made possible to reach subjects scattered throughout the whole national territory, differentiated by title of study, age, gender, hobbies and interests. Compared to the second survey, which totaled 1,400 respondents, the statistics on involvement are as follows: more than 30,000 views, about 2,000 interactions with the post, about 100 reactions (especially likes) and hundreds of comments (as expressions of distrust or annoyance towards the research initiative, real invective in the complaining of the Italian political system and its political representatives, expressions of support towards particular parties/declarations, interactions and sometimes animated discussions among commentators of different political orientations, etc.). As mentioned, the precise evaluation of the first survey results allowed to refine the questionnaire of the second survey, with the goal of being able to compare the data of the two research rounds. For example, we refined the text of questions and answers in the direction of clarification and simplification; sometimes we changed the order of items and replies in case of unreliable data; we eliminated items and response modes in the case of redundancies, of excessively unbalanced data, of distortions linked to social desirability, and also in the direction of thinning and agility of the instrument. All the open/semi-open questions of the first survey were closed (an example for all regards the question about the

motivations related to voting intentions). The accompanying post of the link (through the explicitation of the research theme, of the institutional subjects involved, of the time required for compiling, etc.) has been prepared in a very accurate way. The post has been "placed at the top" within the Department page for the entire duration of the research, the same time span in which the team has constantly monitored the input data and motivated individuals to participate. The module we prepared for the completion of the questionnaire has graphic characteristics that make it aesthetically pleasing. The number of questions is not excessive. There are numerous indications for the correct filling of the questions. Moreover, the technology in use intervenes in order to reset the different erroneous and partial forms of compilation, that are not containable in the case of the paper questionnaires completion: this makes the online completion, for many ways, simpler and more fruitful (few errors and few missing data) than the offline one.

4. The advantages and disadvantages of online data collection

In considering the advantages of a Web survey (Lombi, 2015) we can count on the substantial containment of the research costs (Groves, 1989) on several levels: the sampling tools, the minor human resources employed, the saving of printing questionnaires/doing phone calls, the avoidance of a wide territory mobility, etc. The Internet also has great potential to spread the data collection instrument in a wide and variegated territorial context. The subjects involved in the survey are automatically listed as records of a data matrix. In other words, we have an immediate availability of data for subsequent analyses. We have also already referred to the greater accuracy of the data collected (absence of errors of insertion, absence of errors of compilation, containment of the missing values), as well as their immediate availability in matrix and the possibility of monitoring the results *in itinere*. Obviously, there are also disadvantages: the first one is the statistical non-representativeness of the samples, due to the fact that not all the population uses the Internet and Facebook, while, the second one regards the mechanisms of autoselection and the effect "ball of snow" triggered by the sharing system. The borders of the universe of Facebook subscribers are not defined and, moreover, are constantly evolving; in addition, there is no coincidence between the Social network world and the respective universe of voters present in a given territory, without taking account of the system of basic features of Facebook users, that are not known in a precise way. Finally, we cannot avoid multiple completions made by a unique subject, although we have the possibility to identify and correct them. In fact, in relation to the study on voting intentions in the electoral campaign phase, there are a series of distortions: those due to the applied techniques (standardized questionnaire published on the net) and some other distortions related to the topic (let's think to distrust and resentment towards politics in this historical moment, or to the distrust of some voters, or again to the need for privacy on the personal intentions of voting, etc.).

In particular, in thinking of a negative combination of "technical effect" and "theme effect", there are electoral targets difficult to contact, sometimes impossible to reach. These are the elderly, the non-internet users, the *right* and extremists voters (who show particular distrust, specific cultural and valuable characteristics and sometimes are afraid to express themselves on ideas and practices that they consider to be wholly personal), people with low degree of study, foreigners. On the other hand, we can observe young people, subjects with a high level of education, *left* voters, people who are interested in politics and rather well-informed (it emerges a highly motivated "respondent-type", that is sensitive to research initiatives and interested in politics). Obviously, some distortions in the sample composition, if contained, can easily be corrected through an appropriate weighting.

5. Concluding notes: Online data collection is not enough

It can be concluded that, without solving every criticality emerged but making the utmost effort in the direction of the quality of the achievable empirical basis, the only assumable perspective is the mixed-method approach (Dillman, Smith, Christian, 2009; Amaturo, Punziano, 2016). In our study we highlight the need to combine an offline data collection instrument with one online, taking for granted the same basic characteristics for both (same questionnaire, same mode of administration), in order to address the problem of the coverage of absent or under-represented targets. The email addresses provided for free by a part of the people that we reached online, allowed to proceed with interviews in depth about the anomalous aspects that emerged during the data analysis. The most collaborative subjects have sometimes offered their support to help us in reaching other subjects poorly represented in the sample, with the same characteristics or not, in order to involve them in the investigation. For some aspects, the statistical non-representativeness of the sample gives way to a strong representativeness regarding the substantial plane (Di Franco, 2010). In considering the aim of securing a numerical consistency for each of the electoral targets - known to be diffused in Italy in this historical moment -, we have committed ourselves in this survey (whose publication of the extended version is being set up) to obtain the maximum heterogeneity of socio-demographic and economic variables, despite the distortions generated by the online publication of the questionnaire. All this has been done to be able to accurately investigate values, opinions, perceptions, behaviours in their synergy, with the purpose of identifying electors' styles and profiles who are prevalent and recognizable in our society. The statistical non-representativeness of the sample certainly does not interfere with the theoretical deepening of the theme of political and social participation, neither with the testing and refining of the instruments of data collection, with the activity of conceptualization and operation, with the study of social trends prevailing through the identification of interconnections among variables at multivariate level. On the other hand, a description, a typing and an accurate interpretation of data is a valuable and

fundamental basis also for the preparation of explanatory models and for the identification of predictive factors. All this is based on a deep knowledge of the social and political context of reference, also resulting from the ability to interconnect different and complementary data (for example, analysis of the electoral campaign; analysis of attitudes, voters' intentions and perceptions, aggregate analysis of the outcome of an electoral session, etc.). Mauceri (2003) says about it: "If (...) probabilistic sampling is irreplaceable in research situations in which it is intended to estimate precisely what is the numerical consistency of the diffusion of certain traits within a given population, as when we made opinion polls, the assessment of the relations among variables can require to privilege a comparative logic rather than generalizing, aimed, for example, to compare groups of subjects with opposite action orientations (...) and to establish which are contextual, relational and individual elements that make their courses of action so different".

References

Amaturo, E., Punziano, G., (2016). *I "Mixed-Methods" nella ricerca sociale*, Roma: Carocci.

Boccia Artieri, G., (a cura di), (2015). *Gli effetti sociali del web. Forme della comunicazione metodologie della ricerca sociale*, Milano: Franco Angeli.

Cipriani, R., Cipolla, C., Losacco, G., (a cura di), (2013). *La ricerca qualitativa fra tecniche tradizionali ed e-methods*, Milano: Franco Angeli.

Di Franco, G., (2010). *Il campionamento nelle scienze umane. Teoria e pratica*, Milano: Franco Angeli.

Dillman, D.A., Smyth, J.D., Christian, L.M., (2009). *Internet, mail and mixed-mode surveys. The tailored design method*, San Francisco: Jossey-Bass.

Faggiano, M.P., (2007), *La formazione sociologica nell'università della riforma. La domanda, i percorsi, l'offerta presso il CdL in Sociologia della Sapienza di Roma,* in Fasanella, A. (a cura di), (2007), Milano: Franco Angeli.

Groves, R.M., (1989), *Survey errors and survey costs*, New York: Wiley and Sons.

Kott, P.S., Chang, T., (2010). Using calibration weighting to adjust for nonignorable unit nonresponse, *Journal of the Americn Statistical Association*, 105 (491): 1265-1275, DOI: 10.1198/jasa.2010.tm09016.

Lombi, L., (2015). *Le wen survey*, Milano: Franco Angeli. *Gli effetti sociali del web. Forme della comunicazione e metodologie della ricerca online*, Milano: Franco Angeli.

Mauceri, S., (2003). *Per la qualità del dato nella ricerca sociale. Strategie di progettazione e conduzione dell'intervista con questionario*, Milano: Franco Angeli.

Peruzzi, G., (a cura di), (2017), *Le reti del Terzo Settore. Terzo Rapporto*, Roma: Forum Nazionale del Terzo Settore.

Vaccari *et al.*, (2013), A survey of Twitter users during the 2013 Italian general election, *Rivista Italiana di Scienza politica*, Anno XLIII, n. 3: 381-410, DOI: 10.1426/75245.