#Step @by #Step: recommendations for the development of high quality online research

Otávio Bandeira De Lamônica Freire
Professor da Linha de Marketing do Programa de Pós-Graduação em Administração da Universidade Nove de Julho - PPGA/UNINOVE. Professor Doutor MS3 do Curso de Marketing da Escola de Artes, Ciências e Humanidades - EACH/USP e do Programa de Pós-Graduação em Comunicação da Escola de Comunicações e Artes da Universidade de São Paulo - PPGCOM-ECA/USP. Editor-Chefe da Revista Brasileira de Marketing - ReMARKDiretor e Cofounder da ILUMEO MKT & COM, São Paulo, Brasil
otfreire@usp.br

Diego dos Santos Vega Senise,
Doutorando e Mestre em Comunicação pela Escola de Comunicações e Artes da Universidade de São Paulo - ECA/USP. Diretor e Cofounder da ILUMEO MKT & COWalney Barbosa dos Reis.Bacharel em Marketing pela Escola de Artes, Ciências e Humanidades da Universidade de São Paulo - EACH/USP. Consultor de Data Sciente e Pesquisa de Marketing da Accenture, São Paulo, Brasil
senise@ilumeobrasil.com.br

Walney Barbosa dos Reis
Bacharel em Marketing pela Escola de Artes, Ciências e Humanidades da Universidade de São Paulo - EACH/USP. Consultor de Data Sciente e Pesquisa de Marketing da Accenture, São Paulo, Brasil
walneywebusp@gmail.com

Heder Seiti Ono
Bacharel em Marketing pela Escola de Artes, Ciências e Humanidades da Universidade de São Paulo - EACH/USP. Coordenador Comercial na ILUMEO MKT & COM, São Paulo, Brasil
heder@ilumeobrasil.com.br

Abstract

The present study aims to present and discuss the main concerns over how to ensure the quality of online research. We conducted a state-of-the-art literature review. The findings show the advantages and disadvantages of both online and offline field research, identifying no superiority of one technique over the other. The relevant findings and best practices are presented in the following categories: research project, sample representativeness, questionnaire design and data quality.

Keywords: online research; online sampling; data quality; questionnaire design.

#Step @by #Step: recomendações para o desenvolvimento de pesquisa online de alta qualidade

Resumo

O presente estudo tem como objetivo discutir e apresentar as principais premissas para garantir a qualidade de pesquisas online. O método utilizado foi a revisão do estado da arte da literatura internacional acerca do tema. As principais conclusões apontam para um conjunto de prós e contras em relação às coletas off-line, demonstrando não haver superioridade de nenhuma das duas técnicas. Os principais aprendizados e melhores práticas são apresentados nas seguintes categorias: concepção da pesquisa, representatividade das amostras, design de questionários e qualidade dos dados.

Palavras-chave: pesquisa online; amostragem online; qualidade de dados; design de questionários.

#Step @by #Step: recomendaciones para el desarrollo de encuestas online de alta calidad.

Resumen

Este estudio tiene por objetivo hacer una discusión acerca de las principales premisas que garantizan la calidad de las encuestas online. La metodología utilizada es una revisión de todo que hay de más nuevo e importante en estudios internacionales de esta temática. Las principales conclusiones apuntan para un conjunto de ventajas y desventajas de la investigación online cuando se compara con la investigación offline. Al fin de todo, llegamos a la conclusión de que ningún de los dos tipos demuestra superioridad absoluta. Los aprendizajes más relevantes y las mejores practicas son presentados en estos grupos temáticos: concepción de investigación, muestras y su representatividad, diseño de formularios y calidad de los datos.

Palabras-clave: investigación online; muestra online; calidad de los datos; diseño de formularios.

1 Introduction

The use of online data collection techniques for quantitative research has increased considerably in recent years. Online research has been used by many researchers in Communication, Marketing and Management (Callegaro et al., 2014), especially in Market Research (Comley, 2007), Psychology (Göritz, 2007), Sociology (Tortora, 2008), Electoral Studies (Clarke et al., 2008) and Medicine (Couper, 2007).

ESOMAR[1] estimates that global investment in online research increased from 19% in 2006 to 35% in 2012. Certain features contribute to the success of this technique, such as enabling faster and cheaper data collection for institutes and their clients and optimizing scientific research resources.

The purpose of this article is to unravel the technical and scientific aspects that can ensure quality in all stages of the research process, strengthening the validity of conclusions reached in academia and managerial decisions. Thus, the proposal is to contribute to the development of quality research in the Social Sciences, bolstering the learning achieved in contemporary studies regarding online research. The article also makes recommendations that should be considered in face-to-face (offline) research.

To this end, we undertook a literature review, analyzing the most important articles on the subject in the last fifteen years. The text is divided into four parts: research project, sample representativeness, questionnaire design and data quality (treatment of the response bases).

2 Review Procedures

To achieve the purposes of this work, the following international scientific article databases were searched: Ebsco, Web of Science, Google Scholar and Scopus. The terms used were *online research, online survey, online panel* and *internet-based survey*. Individual searches were also conducted on the websites of publications with a high impact factor and editorial interest in methodological articles in the fields of Marketing, Management, Communication, Public Opinion and Statistics: *Public Opinion Quarterly, International Journal of Market Research, Journal of Survey Statistics and Methodology*, *Journal of Marketing Research*, *Social Science*

---

[1] www.esomar.org. Accessed on 28 July 2016.

Computer Review and *Computers in Human Behavior*. Furthermore, citations were used from articles published in the collection entitled *Online Panel Research: A Data Quality Perspective* (Callegaro et al., 2014).

In total, 8700 articles published in the last 15 years were found containing at least one of the search terms. Filtering by field of interest in the Applied Social Sciences reduced this number to 1200 articles. The final stage of the filtering involved the authors of the present study reading the titles and abstracts to exclude the articles that only applied the online collection technique to test certain theories but were not interested in making methodological discoveries regarding this form of collection. Articles discussing qualitative research were also excluded, maintaining the focus only on quantitative research. In all, 40 articles were analyzed. We concentrated the references on approximately 20 articles, given that the results of the others were similar or complementary to the principal articles. The analysis always took into account the aim of the study, to determine the existence or non-existence of a theory to explain the results, the method used and a discussion of the results. We clustered the findings into 04 categories: research project, sample representativeness, questionnaire design and data quality.

3 Starting from scratch: conception of the research project

The process of conducting research involves a series of processes, knowledge and skills that are acquired through a great deal of study and experience in the field of measurement. This is because conducting research for any purpose is a process with several potential sources of error that need to be duly controlled and understood in detail by the researcher. In other words, it is important to highlight that a "questioner" is not the same thing as a researcher.

One of the major challenges of research in general and, inevitably, online research, is error control. The process of conducting research assumes the possibility of so-called "total error", which is configured as the sum of all the errors that might occur from the conception of the research to the interpretation of the results (Malhotra, 2008). This may be a random sampling or non-sampling error. The errors not related to the sample include non-response, which may occur due to

unavailability of the respondent or refusal to answer. There are also response errors, which involve errors on the part of the researchers, interviewers and interviewees.
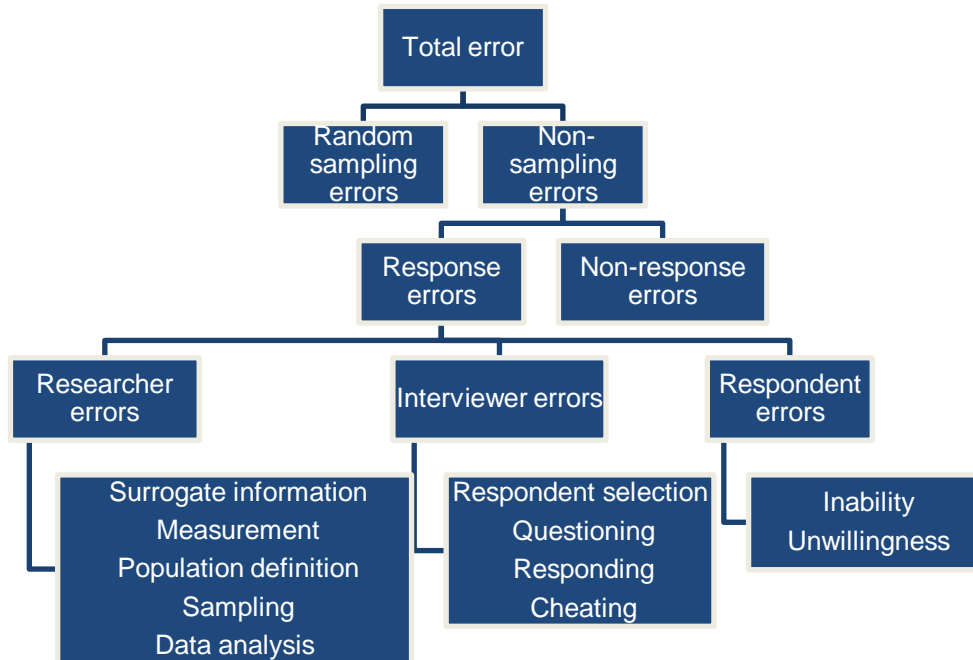

Figure 1
Potential sources of research error
Source: adapted from Malhotra (2008).

Before the research begins, a primordial source of errors that triggers the research process should be considered: the perceived need to measure something. For example, the top executive that needs input for his decision-making process, the university student who wishes to understand a specific human phenomenon, the scientist seeking responses to a theory or the small businessman who wishes to understand why sales in his store are not in accordance with his expectations.

This subject, the "Commissioner of the Research", is the person who sees the need to understand a specific phenomenon. For some reason (prior knowledge, recommendation from colleagues, approach of professional researchers at research institutes, agencies, etc.), he discovers that conducting research may be the solution to his problem.

The Commissioner of the Research may be the researcher himself, but could also be the client that commissioned the research. Therefore, errors accumulate because: (1) the Commissioner of the Research may not know exactly what the

research problem is, (2) the researcher may not know or may not succeed in making the Commissioner of the Research define his needs better and (3) the researcher may not interpret the problem adequately and thus incorrectly define the first decisive step of research: **the research problem**. As stated by Jodorowsky: "Between what I think and what I mean, what I say and what you hear, what you hear and what you think you understood, there is an abyss"[2]. If the researcher has a badly defined problem on his hands, every single measurement (and the entire study) will already be compromised, no matter how sophisticated it is (Sheth & Sisodia, 2002; Butler, 1994).

Therefore, supposing that the research problem is well defined, measurements need to be based on a general research goal, which will require specific objectives and consequent hypotheses. Only with these matters clearly defined can the researcher begin to prepare the questionnaire.

4 Design of online questionnaires

The great advantage of online questionnaires over offline ones lies in their flexibility and considerably greater possibilities for questions and filters, in addition to their distribution on a global scale at a significantly lower proportional cost. Questionnaires with many possible logical pathways (if you answer X, proceed to Question Alpha; if you answer Y, proceed to Question Beta, etc.) are difficult even for professional field researchers to handle. This significantly increases the chance of a tabulation error.

Other evident advantages of online questionnaires are: (1) Instant tabulation: the registered response is automatically added to the database for use in analyses; (2) Simultaneous editing: it is possible to edit the questionnaire rapidly even if it is already being applied if anything wrong is identified or the collection of other data is desired; (3) Minimum costs of reproduction and distribution: traditional questionnaires need to be printed and their cost increases when the sample is large or the quality of the printing is improved; (4) Freedom of the respondent to reflect as long as he deems necessary and respond based on his convictions.

---

[2] http://pensador.uol.com.br/frase/MTU2MDg0MQ/. Accessed on 28 July 2017.

Nevertheless, online questionnaires have certain characteristics that may compromise the quality of the data collected: (1) absence of an interviewer; (2) they are only available to those with internet access (unless the interviewer is available at a location where the respondent is connected to a mobile device) and (3) increased perception of anonymity, which could facilitate and encourage fraudulent responses.

Galesic and Bosnjak (2009) claim that questionnaires designed online tend to generate: (1) responses given in increasingly shorter periods as the questionnaire progresses (2); greater quantity of incomplete responses, (3) shorter responses to open-ended questions and (4) less variability of responses. Sendelbah et al. (2016) emphasize the effect of non-response, especially in a multi-screen and multi-task context, although they contest the reduced variability of responses in evaluation matrices.

To overcome these limitations, it should be remembered that any questionnaire has three specific objectives (Malhotra, 2008):

(1) To transform the desired information into a set of questions that interviewees are capable of answering. Questions that are too long, with too many difficult words or highly complex evaluation matrices tend to reduce the chance of obtaining good quality responses;

(2) To motivate and encourage the interviewee to develop, cooperate and complete the survey. Scrolling designs in mobile devices (advance with the scrollbar of the mouse or cell phone) tend to generate more complete questionnaires, greater quality and variation of responses and fewer technical problems than paging designs (advance with the loading of new pages) in short surveys (Mavletova & Couper, 2014). For longer questionnaires, paging has advantages, as the average time spent on the questionnaire is shorter and there is less satisficing, non-observation of instruction and lower non-response rates (Peytchev, Couper, McCabe, & Crawford, 2006). However, no matter how attractive and easy to respond questionnaires are, a drop in the quality of responses in the last blocks tends to persist. We recommend the randomization of questions and questionnaire blocks, minimizing and distributing this effect; and

(3) To minimize response error. Questionnaires that are very long and not very objective tend to have a higher chance of response error. Another important point is the chosen platform. Some studies require greater attention on the part of the respondent. This may lead the researcher to consider limiting the collection only

to personal computers (supposing that the interviewee uses them in closed and calmer environments, with fewer distractions) or clearly stating that responses need to be recorded in less busy environments and with greater concentration from the respondent, when questionnaires are prepared for mobile devices.

The researcher will now proceed to the sampling and recruitment of respondents, described in the following section.

5 Sampling and recruitment: sample representativeness

Many research companies recruit through online panels. Here, the word panel differs from the traditional use of the term in the field of research. According to the traditional definition, the word is linked to longitudinal studies, research panels that measure the same variables with identical individuals on a number of occasions over time (Hansen, 2008). With the evolution of online research practices, the term panel has come to be used to designate groups of recurring respondents in the digital universe. Millions of people declare that they will cooperate in surveys on various topics at irregular intervals (International Organization for Standardization, 2012, p. 1).

Two basic kinds of online panel make the recruitment, sample selection, types of people interviewed, the interview itself and the types of data to be collected change considerably: probability and nonprobability panels (Callegaro et al., 2014).

6 Probability panels

Probability panels seek the maximum inferential representativeness of a certain population. It is expected that each member of a population of interest, without exception, will have exactly the same probability of being recruited. People are not allowed to participate in a survey unless they have been previously recruited (Calegarro et al., 2014). For example, if the population of interest is Brazil, every Brazilian citizen in the land should have the same chance of being recruited, whether they live in São Paulo or Amazonas, are old or young or have a good or poor financial situation. This is a principle of statistical inference that guarantees, from a

sample, a faithful reflection, with a margin of error and familiar confidence intervals of the given population (Bussab & Morettin, 2010).

Probability panels face two main problems: 1) cost in terms of time and money: either online or using traditional methods, the cost of probability panels with methods such as Random-digit-dialing (RDD) are significantly higher and tend to take longer than nonprobability panels; and 2) availability of the internet in the household, e.g., in Brazil, the National Survey by Home Sampling (PNAD) indicates that only 54.4% of the population over 10 years old accessed the internet at least once in the last 3 months.
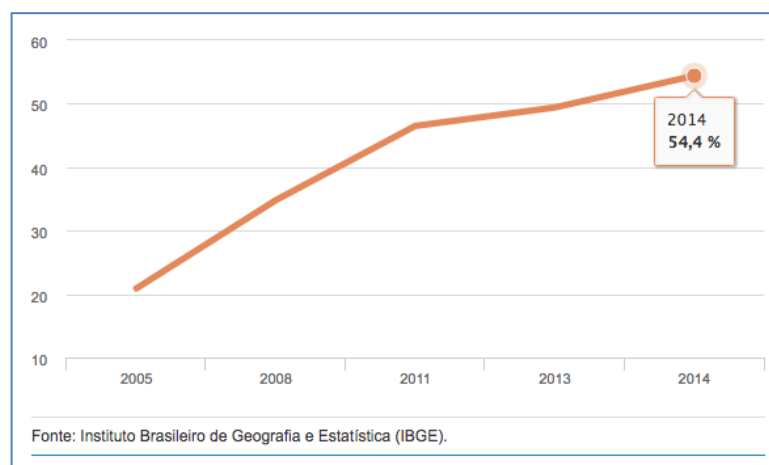

Fonte: Instituto Brasileiro de Geografia e Estatística (IBGE).

Figure 2
Percentage of Brazilian Population with Internet Access[3]
Source: IBGE (2016)

Companies with probability panels use telephone or face-to-face interviews to complete their random sampling. Other companies provide devices with internet access (tablets, cell phones or computers) to selected people in their random sampling (Callegaro et al., 2014).

There are three common types of probability sampling in panels: (1) simple random sampling, (2) quota sampling (defining a maximum number of participants per subgroup according to the research objective) and (3) sample matching.

In 2012, classic random sampling was used by the RAND Continuous Presidential Election Poll (CPEP), considering random selection of census tracts, residences and individuals. Those without internet access were given a WebTV, enabling respondents to access the questionnaire via a television connected to the

---

[3] http://www.valor.com.br/brasil/4513070/mais-da-metade-da-populacao-brasileira-acessa-internet-aponta-ibge Accessed on 20 July 2017.

network. For three months, they all responded weekly regarding their propensity (from 0 to 100) to vote for Obama or Romney. In addition to providing an understanding of changing opinions throughout this period, the method predicted the actual results at the polling booth with only 0.5% error.

An example of sample matching is the Propensity Score Select used by Toluna. The company simultaneously conducts a traditional probability survey and an online nonprobability survey in its panel, with the same questionnaire and sample filters. Following the data collection, the researchers use logistic regression to estimate the probability of an online panel member belonging to the offline group. The technique also allows the variables that increase this probability to be identified. Each respondent in the digital group is given a score for the propensity of belonging to the external (offline) group, and can be recruited for future surveys, maximizing the number of responses within the probability limits (Calegarro et al, 2014).

Global Market Insights uses Pinnacle Methodology (Eggers, 2011). The respondents are classified through 60 demographic, psychographic and behavioral questions used in the US General Social Survey (GSS). From this prior profile of the potential respondents, the samples are designed according to existing distributions in the GSS to increase the representativeness of the sample in surveys.

Probability panels are greatly desirable in terms of quality of the representativeness of the sample and alignment with statistical analysis techniques that assume the probability distribution of the data. However, they involve significantly higher costs and longer data collection periods. Different types of organizations, especially in the private sector, with orientation for rapid decision-making and lower costs, opt for nonprobability (or convenience) panels. In the following section, we provide a definition of a nonprobability panel and present ways of overcoming their limitations in comparison with probability methods.

7 Nonprobability panels

A nonprobability panel is defined by the fact that anyone connected to the internet and somehow exposed to online research can complete available questionnaires, even without prior selection. Irrespective of whether they are part of

the target audience, individuals are free to complete a questionnaire, even if they are excluded by its "filters" (eliminatory questions that classify the respondent).

In a nonprobability panel, it is impossible to know *a priori* all the people invited to take part in the research and the probability of each person participating in the study. In other words, in nonprobability panels, it is not the researcher that chooses exactly who will be included by some previously defined sampling criterion. Thus, the researcher has some limitations in terms of statistical analyses, which have random selection as a basic assumption.

According to Callegaro et al. (2014), the main recruitment methods for nonprobability panels are:

(1)   Purchase of space for publicity online: banners, advertisements on social networks and sponsored links that lead interested parties to register on the websites of the companies that managed panels;

(2)   Invitations to participate on the panel via listing in online or offline groups: a company representative posts to online groups or contacts offline groups (clubs, debating societies, etc.), inviting them to participate in the online penal;

(3)   Co-registration agreements: participants registered with other online services (e-commerce, news sites, etc.) sign terms of commitment agreeing to participate simultaneously in a determined research panel;

(4)   Panel portals or consolidators: companies that allow a user to register automatically for several panels by registering on their website or other media;

(5)   At the end of a survey: when completing a survey, the respondent accessing it for the first time is invited to participate permanently in the panel, leaving his online data for future recruitment; and

(6)   Member-get-a-member (or snowballing): registered members of a panel company distribute surveys to unregistered friends and family in exchange for incentives.

Knowing the methodological limitations, but also the clear advantages in terms of cost and collection time of nonprobability surveys, we can use an array of bias reduction techniques that tend to justify their use for a wide range of purposes.

The use of auxiliary variables is among the most common bias reduction techniques (Calegarro et al., 2014). For instance, if the aim is to represent a population that owns at least one car in a specific region, and 70% of the residents are known to own a car (data that can be obtained through official public channels or

other secondary sources), it is expected that around 30% will not pass the filter of the questionnaire. In other words, around 70% of the respondents in the database have at least one car (if the questionnaire has no filter).

For this type of bias reduction technique to succeed, four elementary criteria must be respected:

(1)   Auxiliary (or comparison) variables should be measured in the questionnaire (Bethlehem & Biffignandi, 2012; Särndal, C. E., & Lundström, 2005);

(2)   The distribution of these variables in the population of interest should be known (Bethlehem & Biffignandi, 2012; Sarndal & Lundstrom, 2005);

(3)   The variables need to be correlated with all the measurements of interest (in other words, the variations in the sample and the population need to be "going in the same direction") (Calegarro et al, 2014); and

(4)   The variables need to be correlated with the probability of response attributed to the typical respondent from the population (in other words, the incidence of certain types of response, from the most to the least common, should correspond to the probabilities extracted from the population of interest) (Bethlehem, Cobben, & Schouten, 2011)

Unfortunately, it is not always possible to obtain important reference variables that go beyond the typical "age", "gender", "education", "city", "income" and others easily found in secondary data. To meet the need for psychographic or attitudinal variables, some companies and institutes use Reference Surveys, which are also discussed in the literature as alternatives for bias reduction (Valliant & Dever, 2011).

Reference Surveys are data surveys that maximize the probability of any individual responding in an area of interest, e.g., the population of Brazil, and collect data with variables to be used as a reference in other studies. For instance, Ilumeo sends a "Welcome" survey to each new member of its online panel[4] with a Member-get-a-member mechanism. In this survey, questions are asked about habits pertaining to the consumption of certain products, media use and other psychographic issues important to its customers, to paint an approximate portrait of the Brazilian population.

The Member-get-a-member mechanism reduces bias by obtaining responses form people with the desired profile, but are not part of the group of people registered

---

[4] http://www.ilumeobrasil.com.br/pesquisas-online.htm. Accessed on 15 August 2017.

in the panel. Other mechanisms that involve purchasing advertising space online (display on websites, search engines or advertisements on social media) also tend to reduce sample inference bias, as it is possible to delimit and proportionally distribute the effort of divulging the survey through highly detailed segmentation of the population of interest. For example, on Facebook, in addition to typical variables such as age, gender and location, it is also possible to segment potential respondents by their behavior in the network (searches for certain products and services). Other options include connection patterns, friends and behavioral demonstrations of interest, such as liking a certain company's page), interests (healthy food, financial investments, vegetarian restaurants, etc.) and, more recently, the possibility of segmentation by income in partnership with Serasa Experian[5].

Other initiatives, such as the ISO 20362:2009 (Access Panels in Market Opinion and Social Research Panel) also define specific quality criteria for panels. This initiative established the double opt-in parameter, where interested people must give clear consent twice. This regulation states that an active member of a panel is only one who has participated in a survey or updated his profile in the last twelve months.

Although nonprobability studies have methodological limitations, it is possible to surround them with bias reduction techniques that justify the balance between minimal loss of quality with an excellent cost benefit, and good response returns in a short time.

8 Incentives to participate in research

In addition to the errors mentioned, two types of error can be minimized with incentives and techniques to strengthen the interviewee's commitment to the survey: non-response errors and response errors (Malhotra, 2008).

Non-response errors emerge due to (1) the absence or unavailability of the respondent when the invitation is made (for example, an online survey banner that was never seen by the potential respondent) or (2) an outright refusal (the respondent is invited, but refuses to participate). In the latter case, incentives tend to increase the probability of a response, as people see a certain value in the difference

---

[5] http://blog.a4d.com.br/facebook-e-serasa-experian-o-que-voce-precisa-saber/. Accessed on 28 June 2017.

between making a sacrifice (participating in the survey) and the benefit of the incentive (cash, points, gifts or taking part in a draw).

Response errors occur for two reasons: (1) inability or (2) unwillingness. Regarding inability (illiteracy, difficulty in using computers, vision problems, etc.), little can be offered in terms of incentives. However, this can be resolved with better delimitation of the public or other collection techniques (letting a person who is close to the respondent and more experienced with computers help elderly or illiterate people, for instance). Unwillingness can be overcome with incentives in an attempt to strengthen an interviewee's commitment to the survey. Punishment (a kind of "reverse incentive") can also be used. For example, if the participant gives clearly incoherent or incomplete responses due to unwillingness, in addition to elimination from the analyses, the panelist that responds unwillingly can be punished by not receiving the promised incentive or permanent exclusion from the panel.

The literature on incentives to participate in surveys has been extensive and consistent over the years. Incentives paid before the research is conducted are more effective in increasing response rates than incentives of the same value awarded later (Lavrakas et al., 2010). Payments can be made in cash, points that can be exchanged for services or the chance to win prizes. Despite the perception that the most effective method is a cash payment, these studies were conducted only with first-time respondents, meaning that this learning cannot be generalized in the long term.

There are mechanisms that increase the respondent's level of commitment[6]. Instead of panelists (people who always complete the questionnaires), it is possible to work with Research Activators, who act as hubs to disseminate online survey links to their social circles. Activators never know the objective of the survey or its filters. If the final respondent (person in the activator's contact network) provides a complete and valid response, the activator is rewarded in cash. Thus, people tend to complete questionnaires more attentively and seriously, as they are motivated to help a friend rather than the institute. By reducing the social distance, the concrete task of responding gains importance.

In the case of panels that do not adhere to the member-get-a-member premise, one of the major problems for researchers when giving incentives for answers is the

---

[6] http://www.ilumeobrasil.com.br/pesquisas-online.htm. Accessed on 21 August 2017.

creation of so-called "professional respondents" (Baker et al., 2013). Professional respondents are people who maintain a network or accumulate benefits by being registered in numerous panels and completing surveys more frequently compared with the typical randomly selected respondent.

The learning accumulated in completing surveys in exchange for incentives can lead to questionnaires with inaccurate and even fraudulent responses. For instance, a professional respondent may lie to avoid filters that he would not avoid if he responded truthfully only to guarantee his reward for completing the questionnaire. Professional respondents also tend not to pay close attention to the questions asked, completing questionnaires in record time, without counting their tendency for satisficing behavior (Krosnick, 1991). This occurs due to the use of "cognitive shortcuts" to provide optimized responses to questions, resorting to responses such as "I don't know how to answer that", socially acceptable responses, acquiescence (agreeing with everything), selecting the first accessible and visibly explicit random selection of responses, ignoring non-compulsory responses or giving open responses with very few words.

An effective technique for avoiding respondents with this type of bias is the use of theoretical constructs with a causal relationship, in addition to analysis of variance and behavior of the data. When the responses are given randomly or with broad agreement or disagreement, the tendency is for there to be no significance in the hypothesis tests related to the causality between the constructs. For example, the greater a person's environmental concern, the greater their intention to purchase green products (Silva, Lima Filho, & Freire, 2015), or the more congruent the celebrity who endorses a brand is, the better the overall evaluation of the advertisement (Aureliano-Silva, Lopes, Freire, & Silva, 2015). When the variables of these two constructs (environmental concern and purchase intention, or celebrity congruence and evaluation of an advertisement) are randomly implemented in a study, the hypothesis test should be significant, with good predictive power (R^2 medium/strong). If it is not, it is highly likely that the set of data has many subjects who responded with satisficing behavior.

To avoid the proliferation of low quality responses by professional respondents, panel companies have adopted several procedures (Baker et al., 2013), including:

(1)   Trap questions: these do not generally improve the analyses of the study, but they can identify less attentive respondents (Oppenheimer et al., 2009).

For instance, the respondent is asked to evaluate X brand of soap powder. At the end of the questionnaire, a question emerges (such as "Which brand of soap powder did you evaluate at the beginning of the questionnaire?") to eliminate questionnaires with inadequate responses;

(2)   Search for false identities: removing panelists who create more than one account with other identities to complete the same questionnaires more times and, consequently, earn more;

(3)   Collecting more responses than necessary: the goal here is to have a margin for removing bad or incoherent responses, questionnaires completed quicker than expected or identified as being from professional respondents. The number of surplus responses to be collected may vary according to the general quality of the panel in question. However, Downes-Le Guin et al. (2006) suggest that between 1% and 5% of collected responses are fraudulent and are worthless to the researcher. In other words, this is the margin of excess that can be maintained to ensure the removal of responses without affecting the final sample size;

(4)   Techniques for obtaining responses and constant updating of the respondent base: knowing that there is a tendency towards professionalization as the number of questionnaires responded increases, an alternative is the constant updating of the panelists and intelligent selection of respondents. All the surveys available do not necessarily have to be made available to every registered panelist. The average number of questionnaires per respondent tends to fall, increasing their "useful life" as non-professional respondents. Nevertheless, if few surveys fit a panelist's profile, the dropout rate can also increase. With the member-get-a-member method, there are lower rates of professional respondents, because the probability of reaching new respondents tends to increase; and

(5)   There must be a balance between the level of difficulty of the questionnaire and the incentive offered: very attractive incentives tend to arouse great interest among professional respondents increasing the rate of fraudulent responses. Very weak incentives are also not desirable, as the respondent sees no benefit in beginning the questionnaire or in completing it adequately. The balance between the difficulty of the questionnaire and the benefit offered can be determined by using A/B tests or more sophisticated techniques, such as Conjoint Analysis (Green & Srinivasan, 1978; Rao & Pilli, 2014), which will provide a better combination

of incentives, number of questions, average execution time and even the research theme.

After the data collection, we reached the data analysis phase, with key questions regarding the quality of the treatment of the responses and the interpretation of the results.

9 Data Quality: Treatment of responses

In 2007, the Advertising Research Foundation (ARF) created the Online Research Quality Council (ORQC). One of its first goals was to identify the potential and limitations of the 17 suppliers of online panels in the USA through comparative studies. They all used nonprobability collection methods. A two-wave study was conducted from October to November 2008. The same questionnaire was forwarded to participants from all the panels, administered by an exempt company that was not involved in the study. The questionnaire contained attitudinal (continuous) and behavioral (categorical) questions. In total, there were 1,038,616 invitations and 76,310 responses. As well as being compared with each other (17 online panels), the results were also compared with benchmarks (in this case, reference metrics), probability surveys conducted by the government.

The study (Walker et al., 2009) found significant differences between the results of the panels and the benchmarks. In general, the panels overestimated the behavioral results. For example, on the rate of smokers in the population (people who had smoked at least 100 cigarettes in their lives), some panels estimated this number at 58%, while others were nearer to the benchmark of the NHIS (National Health Interview Survey), at 42%. It is necessary to consider that online surveys increase the perception of the respondent's anonymity, reducing the number of socially desirable responses. It is possible that some people would be more willing to "admit" that they smoke when responding online rather than in person.

In other less compromising behavioral questions, the difference between the results of the panels and the benchmarks was less striking. For instance, regarding ownership of cell phones, the variation between the panels was 85% to 93%, with the probability benchmark indicating 79%.

In addition to comparison with the benchmarks, it is important to highlight the significant differences of results from one panel to another. The most contrasting

results between the panels were found in questions related to consumers' attitudes. The intention to purchase a brand of soup was tested. Analyzing the frequency of responses with the two most positive alternatives (top 2 boxes), the variation between the panels was high. The panel with the lowest result indicated 32% acceptance of the product, whereas the panel with the highest rate of acceptance indicated 53%. This indicates that online panels are not similar to one another. The quality and accuracy of the final results depend on the technical protocols the companies use to attract and manage their respondents, as well as the treatment of their databases.

Part of the problem lies in the sampling process and respondent selection diagnosed earlier and with control alternatives. From the data collection, Ilumeo, for example, uses criteria for cleaning the base known as IQC (Ilumeo Quality Criteria). These criteria involve:

1)   IP (Internet protocol) tracking and geolocation: the respondent's IP is traced and consequently more than one response from each connected device is not permitted. Only in very rare cases is more than one IP permitted (telemarketing operations and other companies with hundreds of computers connected to a network, generating a single IP, for instance). Furthermore, the IP allows the respondent's location to be tracked, flagging incoherencies regarding the respondent's declared place of residence;

2)   Response time: response times well below the average are immediately removed, as are those well above the average time in both the questionnaire and per page (Malhotra, 2008; Greszki et al., 2015);

3)   Incoherence: trap questions and other forms of incoherent responses are tracked and eliminated;

4)   Univariate outliers: people who do not vary their responses, which may indicate a lack of commitment to the study (Levin, Fox, Forde, & David, 2012);

5)   Multivariate outliers: extremely varied responses, but without a pattern, theoretical and conceptual support or strange atypical responses are also eliminated (Levin et al., 2012);

6)   Response rate: high response rates must be analyzed. Very high incentives, as mentioned earlier, tend to be the main factors responsible for this. A

study conducted by Yeager et al. (2011) suggests that high response rates are generally associated with low quality data;

7)   Variance test for Common Method Bias: in questionnaires applied in a single stage, it is recommended that an analysis should be conducted regarding whether the data behavior does not merely reflect the bias generated by the questionnaire itself (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003);

8)   Test of difference between waves: The respondents are divided into groups per quartile of response time to evaluate whether there is a significant difference in the dependent variables between groups (Zaller & Feldman, 1992); and

9)   Test of Relationship between Constructs: as mentioned above, in the case of broad theoretical predictability of the relationship between constructs, this relationship should also be significant in the sample set.

10 Data analysis and interpretation of results

In procedures for cleaning the response base, we come to data analysis. As it is a complex subject involving diverse techniques such as descriptive analyses, different averages, correlation, regression, factor analysis, variance and covariance analysis, discriminant analysis, multidimensional scaling, joint analysis, Logit analysis or structural equations, we will end the block by asking: Did all these analyses answer the research problem set by the Commissioner of the Research?

If the initial problem were "Why do people stop buying a product?", the analysis should generate either: (1) conclusions (in case of experimental or conclusive research); or (2) well-based assumptions (based on exploratory research). An analysis of this problem would normally generate answers such as, "because the profile of the public has changed", "because the product was replaced by an equivalent or better one", "because there are problems in the distribution channels, such as a shortage of the product in stock (rupture)" or "because there is a new competitor with a lower price and similar quality".

The Table 1 summarizes the main findings of the articles that made the greatest contributions to this review, presenting the main discoveries of each study and the broader theme that guided the development of the present article:

Table 1

Summary of the main findings of the analyzed articles

Author	Journal	Discovery	Theme
Scherpenzeel & Toepoel (2012)	Public Opinion Quarterly	The more effective response rate was due to prior incentives of 10 euros. Incentives of 20 to 50 euros did not substantially increase the response rates beyond those achieved with 10 euros. Prior incentives are interpreted as a vote of confidence in the respondent. They trigger a sense of obligation to respond in keeping with the norm of reciprocity.	Research Project
Lugtig &Toepoel (2016)	Social Science Computer Review	There is no change in the pattern of measurement errors when an individual responds in different devices (smartphones, tablets or computers). It is estimated that greater error is attributed to tablets because the device is chosen by the respondent himself.	Questionnaire Design
Mavletova & Couper (2014)	Journal of Survey Statistics and Methodology	Short questionnaires for mobile devices (17 questions) with the respondent using the scrollbar leads to quicker completion, lower (albeit not significant) rates of interruption, fewer technical problems and attitudinal responses with greater variance, as well as greater respondent satisfaction with the process. The sending of text messages (rather than e-mails) significantly increases participation rates with mobile devices.	Questionnaire Design
Sikkel et al. (2014)	Public Opinion Quarterly	For most formats, "dragging" rather than clicking takes up more of the respondent's time. However, dragging has a positive effect on satisfaction with the survey. Respondents feel that the questionnaire is more pleasant, interesting and relevant to them, even if they find it takes longer. Nevertheless, it is a design that is more pleasant at first contact. When repeated, the vividness of the procedure does not compensate for the extra time required to complete it.	Questionnaire Design
Mavletova (2013)	Social Science Computer Review	Full response rates on mobile devices were fewer, as was the size of open responses.	Questionnaire Design
Peytchev et al. (2006)	International Journal of Public Opinion Quarterly	Extensive survey (268 questions) with 21,000 undergraduate students. Respondents who used paging in general had a shorter response time, less satisficing behavior, observed instructions more closely and had lower non-response rates.	Questionnaire Design
Eckman (2016)	Social Science Computer Review	The inclusion of residences that do not regularly use the internet showed substantial socio-demographic differences in relation to those that use it regularly. However, there were few	Sample Representativeness

		significant differences in the estimates of the models tested in the 5 studies with sampling in the same panel.	
Blom et al. (2016)	Social Science Computer Review	The exclusion of "offline households" produces a significant bias in the coverage by the panels, while their inclusion improves the representativeness of the sample, despite the low propensity to complete the questionnaires.	Sample Representativeness
Yeager et al. (2011)	Public Opinion Quarterly	(1) Surveys with representative random samples were consistently more accurate than non-representative ones, even after post-stratification. However, post-stratification improved the accuracy of some non-representative samples. (2) Participation and response rates were negatively related to accuracy, challenging the notion that high conversion and response rates are indicators of data consistency. Accuracy was analyzed by comparing the estimates of surveys with official government benchmarks or high quality federal research with high response rates.	Sample Representativeness
Ansolabehere; Schaffner (2015)	Journal of Survey Statistics and Methodology	Distractions are common and significantly increase the duration of surveys. However, there is little evidence for a change in the quality of responses. Likewise, interruptions did not affect whether responses to factual questions were right or wrong.	Data Quality
Greszki et al. (2015)	Public Opinion Quarterly	The article explores the response time of so-called speeders (people who complete questionnaires quickly), using the response processing model. It is usually argued that this type of respondent does not pay due attention and thus the quality of their responses is considered poorer. Nevertheless, an analysis of 9 online surveys indicates that this occurs more at the page-wise level (time spent on a page) than the general response time (case-wise), considering two standard deviations.	Data Quality
Toepoel et al. (2008)	Public Opinion Quarterly	Respondents with greater experience on panels complete questionnaires more quickly and with greater correlations. They also tend to choose the first option shown, suggesting that more experience on panels leads to less commitment to the survey in question before completing it.	Data Quality
Malhotra (2008)	Public Opinion Quarterly	Participants with lower levels of education, and who complete questionnaires more quickly, have a greater propensity to effects of primacy in "unipolar" scale items.	Data Quality
Sendelbah	Computers in	In multi-screen and multi-task contexts,	Data Quality

| (2016) | Human Behavior | there is a greater tendency to non-response, but not necessarily a reduction in the variability of responses. | |

Source: elaborated by the authors (2017)

11 Final Considerations

The purpose of this study was to survey the state-of-the-art in the development and operationalization of assertive and reliable online research projects from the viewpoint of the general quality of responses and collected data. The authors opted to examine the literature of the last 15 years in journals in the fields of Marketing, Management, Communication, Public Opinion and Statistics. The articles that were analyzed contributed to their fields in terms of methodology and operation of research and were clustered into four major themes: research project, sample representativeness, questionnaire design and data quality.

The results infer that there is no specific stage of the project that deserves more or less attention and dedication from all those involved in the development of the project, operationalization in the field and preparation of the data for analysis and recommendations (designer, researcher, interviewer, respondent, analyst who treats the data). Errors can occur in any of the stages.

This article contributes to the development of quality research in the field of Social Sciences as it includes diverse forms of learning from contemporary studies on online research. It also makes recommendations to be considered in face-to-face (offline) studies. It has a sequence, a framework that ensures the general quality of responses and the data to be used. This framework should be used by academic and market researchers, given that the intrinsic goals of these "two worlds" are bound to overlap in the sense that it is necessary to ensure that interpretations will not be biased due to errors in design, the field and the treatment of collected data.

Regarding the quality and reliability of online survey data, it can be concluded that online research has an excellent cost-benefit and is high in quality. Indeed, the definition of the problem, the questionnaire design, sampling method and quality controls of the databases are much more relevant than the pure and simple answer to the question "Is it better to do an online or traditional survey? There is no perfect option, and no consensus regarding which is the best approach. This means that the

reliability of online methods cannot be judged appropriately in comparison with traditional research (Kellner, 2004). In other words, both techniques have their pros and cons, and it falls to researchers to choose the one that is best suited to their purposes.

References

Advertising Research Foundation. (2013). Online Survey Research: Findings, best practices, and future research. *Research Quality Forum.*

Ansolabehere, S., & Schaffner, B. F. (2015). Distractions: The Incidence and Consequences of Interruptions for Survey Respondents. *Journal of Survey Statistics and Methodology, 3*(2), 216-239.

Aureliano-Silva, L., Lopes, E. L., Bandeira De Lamônica Freire, O., & da Silva, D. (2015). The brand's effect on the evaluation of advertising endorsed by celebrities: an experimental study. *BBR-Brazilian Business Review, 12*(4).

Baker, R., Brick, J. M., Bates, N. A., Battaglia, M., Couper, M. P., Dever, J. A., & Tourangeau, R. (2013). Summary report of the AAPOR task force on non-probability sampling. *Journal of Survey Statistics and Methodology, 1*(2), 90-143.

Bethlehem, J., & Biffignandi, S. (2011). *Handbook of web surveys* (Vol. 567). John Wiley & Sons.

Bethlehem, J., Cobben, F., & Schouten, B. (2011). *Handbook of nonresponse in household surveys* (Vol. 568). John Wiley & Sons.

Blom, A. G., Bosnjak, M., Cornilleau, A., Cousteaux, A. S., Das, M., Douhou, S., & Krieger, U. (2016). A comparison of four probability-based online and mixed-mode panels in Europe. *Social Science Computer Review, 34*(1), 8-25.

Bussab, W. D. O., & Morettin, P. A. (2010). *Estatística básica*. Saraiva.

Butler, P. (1994). Marketing Problem: From Analysis to Decision. *Marketing Intelligence & Planning, 12*(2), 4-12.

Callegaro, M., Baker, R. P., Bethlehem, J., Göritz, A. S., Krosnick, J. A., & Lavrakas, P. J. (Eds.). (2014). *Online panel research: A data quality perspective*. John Wiley & Sons.

Clarke, H. D., Sanders, D., Stewart, M. C., & Whiteley, P. (2008). Internet surveys and national election studies: A symposium. *Journal of Elections, Public Opinion and Parties, 18*(4), 327-330.

Comley, P. (2007). Online Market Research. *Market Research Handbook* (5a ed.). Chichester: John Wiley & Sons.

Couper, M. P. (2007). Issues of representation in eHealth research (with a focus on web surveys). *American journal of preventive medicine*, *32*(5), S83-S89.

Toepoel, V., Das, M., & Van Soest, A. (2008). Effects of design in web surveys: Comparing trained and fresh respondents. *Public Opinion Quarterly, 72*(5), 985-1007.

Downes-Le Guin, T., Mechling, J., & Baker, R. (2006). Great results from ambiguous sources: Cleaning Internet panel data. *ESOMAR World Research Conference: Panel Research*.

Eckman, S. (2016). Does the Inclusion of Non-Internet Households in a Web Panel Reduce Coverage Bias? *Social Science Computer Review, 34*(1), 41-58.

Eggers, M., Drake, E. (2013). Blend, balance and stabilize respondent sources. *75th Annual conference of the Advertising Research Foundation, 1*. 2013.

Galesic, M., & Bosnjak, M. (2009). Effects of questionnaire length on participation and indicators of response quality in a web survey. *Public opinion quarterly, 73*(2), 349-360

Green, P. E., & Srinivasan, V. (1978). Conjoint analysis in consumer research: issues and outlook. *Journal of consumer research, 5*(2), 103-123.

Göritz, A. S. (2007). Using online panels in psychological research. *The Oxford handbook of Internet psychology*, 473-485.

Greszki, R., Meyer, M., & Schoen, H. (2015). Exploring the effects of removing "too fast" responses and respondents from web surveys. *Public Opinion Quarterly, 79*(2), 471-503.

Hansen, J. (2008). Panel surveys. *The Sage handbook of public opinion research*, 330-339.

International Organization for Standartization. (2012). ISO 20252: *Market, opinion and social research: vocabulary and service requirements.* Geneva.

International Organization for Standartization. (2009). ISO 26362: *Market, opinion and social research: vocabulary and service requirements*. Geneva: ISO.

Kellner, P. (2004). Can online polls produce accurate findings? *International Journal of Market Research, 46*(1), 3-22.

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied cognitive psychology, 5*(3), 213-236.

Lavrakas, P., Dennis, M., Disogra, C., Subias, T., Lawarence, M., & Tompson, T. (2010). Panel conditioning and attrition in the AP-Yahoo News election panel study. *Proceedings of the Joint Statistical Meeting.* Washington.

Levin, J. Fox, J., Forde, A., & David, R. (2012). *Estatística para Ciências Humanas*.

Lugtig, P., & Toepoel, V. (2016). The use of PCs, smartphones, and tablets in a probability-based panel survey: Effects on survey measurement error. *Social Science Computer Review*, *34*(1), 78-94.

Malhotra, N. (2008). Completion time and response order effects in web surveys. *Public Opinion Quarterly*, *72*(5), 914-934.

Mavletova, A. (2013). Data quality in PC and mobile web surveys. *Social Science Computer Review*, *31*(6), 725-743.

Mavletova, A., & Couper, M. P. (2014). Mobile web survey design: scrolling versus paging, SMS versus e-mail invitations. *Journal of Survey Statistics and Methodology*, *2*(4), 498-518.

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, *45*(4), 867-872.

Peytchev, A., Couper, M. P., McCabe, S. E., & Crawford, S. D. (2006). Web survey design: Paging versus scrolling. *International Journal of Public Opinion Quarterly*, *70*(4), 596-607.

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of applied psychology*, *88*(5), 879.

Rao, V. R., & Pilli, L. E. (2014). Conjoint Analysis for Marketing Research in Brazil. *REMark: Revista Brasileira de Marketing*, *13*(4).

Särndal, C. E., & Lundström, S. (2005). *Estimation in surveys with nonresponse.* John Wiley & Sons.

Scherpenzeel, A., & Toepoel, V. (2012). Recruiting a probability sample for an online panel: Effects of contact mode, incentives, and information. *Public opinion quarterly*, *76*(3), 470-490.

Sendelbah, A., Vehovar, V., Slavec, A., & Petrovčič, A. (2016). Investigating respondent multitasking in web surveys using paradata. *Computers in Human Behavior*, *55*, 777-787.

Sheth, J. N., & Sisodia, R. S. (2002). Marketing productivity: issues and analysis. *Journal of Business research*, *55*(5), 349-362.

Sikkel, D., Steenbergen, R., & Gras, S. (2014). Clicking vs. dragging: Different uses of the mouse and their implications for online surveys. *Public opinion quarterly*, *78*(1), 177-190.

Silva, Q. F., Lima Filho, D. O., & Freire, O. (2015). A influência da consciência ambiental e das atitudes em relação ao consumo sustentável na intenção de compra de carne bovina. *Revista de Administração da Universidade Federal de Santa Maria*, *8*(3), 463-481.

Tortora, R. *Recruitment and Retention for a consumer Panel*. (2008). Hobroken, NJ: John Wiley & Sons.

Valliant, R., & Dever, J. A. (2011). Estimating propensity adjustments for volunteer web surveys. *Sociological Methods & Research*, *40*(1), 105-137.

Walker, R., Pettit, R., & Rubinson, J. (2009). *The Foundations of Online Research Quality–Executive Summary 3: Inter Study Comparability and Benchmark Analysis.* New York: Advertising Research Foundation.

Yeager, D. S., Krosnick, J. A., Chang, L., Javitz, H. S., Levendusky, M. S., Simpser, A., & Wang, R. (2011). Comparing the accuracy of RDD telephone surveys and internet surveys conducted with probability and non-probability samples. *Public opinion quarterly*, *75*(4), 709-747.

Zaller, J., & Feldman, S. (1992). A simple theory of the survey response: Answering questions versus revealing preferences. *American journal of political science*, 579-616.