

# Data collection in Internet environment in social representations studies\*

Recolección de datos en Internet en estudios de representaciones sociales

JOAO FERNANDO RECH WACHELKE\*\*

*Università degli studi di Padova, Italy*

BRIGIDO VIZEU CAMARGO

*Universidade Federal de Santa Catarina, Brazil*

SAMUEL LINCOLN BEZERRA LINS

*Universidade Federal da Paraíba, Brazil*

ALINE VIEIRA DE LIMA NUNES

*Universidade Federal da Paraíba, Brazil*

## Abstract

The present study aimed at comparing social representations structures concerning data collection procedures: through internet forms, diffused in the WWW, and through conventional paper and pencil questionnaire methods. Overall 893 individuals participated in the research, 58% of whom were female. A total of 217 questionnaires about the social representation on football (soccer) and 218 about the representation on aging were answered by Brazilian university students in classrooms. Electronic versions of the same instrument were diffused through an internet forum linked to the same university. There were 238 answers for the football questionnaire and 230 for the aging one. The instrument asked participants to indicate five words or expressions related to one of the social objects. Sample characteristics and structural analyses were carried out separately for the two data collection procedures. Data indicated that internet-based research allows for higher sample diversity, but it is essential to guarantee the adoption of measures that can select only desired participants. Results also pointed out the need to take into account the nature of the social object to be investigated through internet research on representations, seeking to avoid

self-selection effects, which can bias results, as it seems to have happened with the football social object.

*Keywords:* social representations, Internet-based research, soccer, aging, Brazil.

## Resumen

El presente estudio buscó comparar las estructuras de representaciones sociales según los procedimientos de recolección de datos: a través de formatos difundidos por Internet y a través de métodos de cuestionario convencional de lápiz y papel. En total, 893 personas participaron en la investigación, 58% de las cuales fueron mujeres. Un total de 217 cuestionarios sobre representaciones sociales sobre el fútbol y 218 sobre envejecimiento fueron respondidos en el aula de clase por estudiantes universitarios brasileños. Las versiones electrónicas del mismo instrumento fueron difundidas a través de un foro en Internet, ligado a la misma universidad. Hubo 238 respuestas para el cuestionario de fútbol y 230 para el de envejecimiento. El instrumento pidió a los participantes indicar cinco palabras o expresiones relacionadas con uno de los objetos sociales. Se efectuaron caracterizaciones de la muestra y análisis estructurales de cada uno

\* The authors would like to thank researchers Andréa Barbará Bousfield and Izabela Sampaio for their comments about a first version of this paper.

\*\* Correspondence: Joao Fernando Rech Wachelke, Università degli studi di Padova, Italy. E-mail: wachelke@yahoo.com.

de los dos procedimientos de recolección de datos, por separado. Los datos mostraron que la investigación basada en Internet permite una diversidad de muestra más grande, pero esto es esencial para garantizar la adopción de medidas que permiten seleccionar únicamente a los participantes deseados. Los resultados también apuntan a la necesidad de tomar en cuenta la naturaleza del objeto social investigado a través de la investigación en Internet sobre representaciones, buscando evitar los efectos de auto-selección, los cuales pueden sesgar los resultados, lo cual parece haber sucedido con el objeto social del fútbol.

*Palabras clave:* representaciones sociales, investigación basada en Internet, fútbol, envejecimiento, Brasil.

### Introduction

With internet diffusion, a growing contingent of researchers have come to carry on studies with participant recruitment and data collection stages based in World Wide Web environments (Birnbaum, 2000). The ease of access to participants, in some cases, and the possibility of reaching populations from online communities that would be hard to contact in conventional conditions are among the most important advantages of conducting studies over the internet. Moreover, there are economic advantages, which allow for a democratization of knowledge production: according to Kaplowitz, Hadlock and Levine (2004), obtaining a response from a participant in research hosted on the internet could cost up to 8 times less than a response to a regular questionnaire. The main reasons that Web experimental research pioneers mentioned to execute studies through the internet were: large samples in internet studies, high statistical power of experiments with those samples and low execution costs (Musch & Reips, 2000).

There are two basic assumptions that must be met to consider internet-based research as a scientifically valid practice. The first one is that psychological processes operate similarly in people using

the internet and people not using it. The second one is that it is possible to have access to diverse, if not representative, samples of the populations of interest (Best, Krueger, Hubbard & Smith, 2001). In general, a correspondence between surveys, correlational studies and experiments carried out in the internet and in a conventional, “paper and pencil” way (Krantz & Dalal, 2000; McGraw, Tew & Williams, 2000; Andrade & Wachelke, 2006; Wachelke & Andrade, 2006).

Concerning samples, it is known that they can be diverse, but not representative in a statistical sense. This problem is called coverage error (Couper, 2000). The population with access to the net has higher education levels than the population in general (Fricker, Galesic, Tourangeau & Yan, 2005). In Brazil, access to the internet is higher to people with high family income, high school level and age between 15 and 24 years (Instituto Brasileiro de Geografia e Estatística,<sup>1</sup> 2006). In those cases, it is necessary to define reference population as people with access to the internet, instead of the general population (Couper, 2000) or to investigate populations that have similar parameters to those of the population with internet access (Hayslett & Wildemuth, 2004). In internet-based research the existence of self-selection is also to be considered (Reips, 2000; 2002), since a given participant has more autonomy to decide which studies to take part on, and net research favors the participation of motivated people. In this case, there is also the risk of covering a group that might have little in common with the population segment of interest from internet outsiders, and taking the risk of restricting itself to a “volunteer psychology”, which tends to occur less in conventional data collection situations (Reips, 2000).

Internet-based research presents itself as a useful tool for social representations studies. Social representations have found new life as a sociopsychological phenomenon with Moscovici's (1961/1976) study about the French people's social representation on psychoanalysis. Considering that social representations are symbolic constructions that are elaborated and shared by groups about social

1 Brazilian Geography and Statistics Institute.

objects, which enable group members to deal with aspects of reality that are relative to those objects (Jodelet, 1989; Abric, 1998; Wagner, 1998), it is possible to carry out surveys or experiments in internet environments with the intending to describe representation fields and contents. Various data analysis procedures referring to social representations are based in large samples, with more than 200 participants, and at times it is not viable to find people in a satisfactory number that could be inserted in this condition, due to the specificity of some of the researched groups.

However, coverage and self-selection issues could compromise social representation studies even more than basic psychological processes studies, since it is essential to guarantee that responses provided by participants recruited through the internet, whether by e-mail messaging to a specific social segment or by means of messages spread in discussion forums, are close from those that would be emitted by participants answering to traditional collection procedures. In other words, do representations reconstituted through a data collection executed in internet environment and through data from conventional paper and pencil questionnaires have the same structure?

The present study aims at verifying if there is empirical convergence of results following those two data collection procedures, as a way of assessing the validities of samples obtained through the Web. The structural perspective on social representations has been adopted (Abric, 1998). According to such approach, what determines a representation's identity is its central core; there are different representations when elements contained in their cores are different.

## Method

### Research design

The study was an experiment with independent measures. There were two independent variables: social representation object and data collection procedure. The social object variable concerned the object of representation about which participants were asked to respond about: football or aging. The

investigation of different social objects is justified by a bigger possibility of generalization of results. The data collection "procedure" variable had two modalities: conventional in-person data collection, with the aid of print questionnaires to be filled with pens or pencils and recruitment in university classrooms through oral invitation; and internet-based data collection, with participant recruitment in internet communities linked to the same university and filling in an electronic form. Thus, there are four experimental groups: participants that answered questionnaires about football or aging divided in the two data collection strategies.

### Participants

A number of 893 took part on the study, 58% of whom were female. Participants' ages ranged from 14 to 62 years old ( $M = 23$  years and 2 months, median: 22 years and  $SD = 5$  years and 7 months).

### Instrument

The research instrument was a questionnaire in print and electronic formats. It asked participants to perform an evocation task, indicating five words or expressions brought to their minds by the inducing words football or aging, according to the experimental group in which they were inserted. There were also participant characteristics' items such as gender, age, and if the participant was or was not a university student. On the internet-diffused form, the participant's Brazilian geographic region of residence was also asked.

Finally, there were items whose goal was to measure participants' proximity with the social objects that were proposed. In the questionnaires about aging, participants should indicate if they had little contact with elderly people, if they had occasional contact with elderly people or if there was one or more elderly person living with them. About football, they should select one of the situations that described their relationship with the sport: did not follow football, only supported a football team, supported a football team and also followed the sport closely, and supported a football team to a level that included going to stadiums to watch matches.

## Procedure

Conventional in-person data collection happened in the second semester of 2006. Two trained research assistants invited undergraduate students that were attending classes at a federal university from the Northeast of Brazil to participate in the study. Students from exact, social and biological sciences were included in the sample. In the classrooms, questionnaires were randomly distributed in terms of designing participants to answer about one of the social objects. Data collection lasted for a week coming to an end after little more than 200 questionnaires for each of the social objects was administered.

On the internet, a message inviting undergraduate students to participate in the study was posted on a site dedicated to students and professors from the same university in which the conventional collection took place. Such site was an online community with more than 7000 individuals subscribed to it. Participation involved answering to a form to which a hyperlink was provided on the message text. The invitation was repeated daily on the site, to guarantee that it was visible among the discussion forum's topics. The form was kept online until little more than 200 individuals had responded for each social object, what took around a month.

Both in the verbal invitation in the conventional data collection modality and in the message posted in the internet there were instructions that made it clear that only students enrolled at the specified university could participate. Overall 217 questionnaires about the social representation on football and 218 about the social representation on aging were answered for the conventional procedure. Internet-based data collection generated 238 responses for the football questionnaire and 230 for the one related to aging.

## Data analysis

The first procedure involved differences analysis between the experimental groups in the participant characteristics variables. The structural analysis for both representations occurred separately for the two experimental conditions concerning data

collection procedure, and employed an adaptation of Vergès (1992) prototypical analysis. Individual responses were thematically categorized, and then categories' mean evocation orders and evocation frequencies were crossed in tables allowing for an evaluation on the hypothetically central elements for each experimental condition.

## Results

An important aspect is the fact that internet samples had participants that were not enrolled to the university in proportions higher than 10%, which shows smaller precision in the access to the population of interest, even though there were explicit instructions on participation restriction. While there were not any non-participants that answered to the conventional situation questionnaires, for the internet conditions there were 36 (football) and 29 (aging).

For the following analyses only the results of internet-based data collection with students were considered. A total of 203 questionnaires about football and 201 on aging were maintained.

For both social objects there was more diversity in participation courses on the internet-based conditions. For the aging social object, there were participants enrolled in 14 courses for the conventional response condition, while 36 courses were represented on internet data collection. For the football social object, those numbers were 16 and 32, respectively.

There were no significant differences in participant characteristics variables for the aging questionnaire. From the participants that answered to the conventional situation questionnaire, 119 (54.6%) were female; for internet participants, 122 (60.7%) ( $\chi^2 = 1.60$ ;  $df = 1$ ; ns). On the conventional response situation, mean participant age was 22 years and 6 months old ( $SD = 4$  years and 10 months), and on the internet condition, it was 23 years and 4 months old ( $SD = 5$  years and 8 months) ( $t = 1.51$ ;  $df = 412$ ; ns). No association was found also between data collection procedure and participant proximity degree with aging ( $\chi^2 = 1.23$ ;  $df = 1$ ; ns). Considering the two conditions together, 132 individuals (30.8%) said to have little contact with

elders, 110 (54.7%) reported occasional contact and only 64 (15.3%) said to live with elders at home.

On the other hand, for the football social object, there was an increase in proportion of male participants on the internet condition ( $\chi^2 = 8.18$ ;  $df = 1$ ;  $p < 0.01$ ), that were a minority on the conventional data collection condition and are practically half of the participants that took part on the internet collection. There was also a different distribution in the intensity levels of supporting practices ( $\chi^2 = 9.66$ ;  $df = 3$ ;  $p < 0.05$ ), indicating a change in profile of participants that responded to football questionnaires. Finally, internet participants were also older ( $t = 2.66$ ;  $df = 408$ ;  $p < 0.01$ ) (see table 1).

Table 2 presents elements located on the four experimental conditions' first quadrants in prototypical analysis. Those are elements with low evocation orders and high frequencies, compared

**Table 1. Participant characteristics variables for participants that answered to the questionnaires about the social representation on football**

	Football - conventional	Football - internet
% Male gender	35.9	49.8
Mean age	21 years 8 months (SD 4.01)	22 years 9 months (SD 3.91)
% Does not follow football	33.3	47.8
% Supports a team	33.6	25.6
% Supports a team and follows	23.0	19.7
% Supports a team and stadium	10.1	6.9

Fuente: autores.

**Table 2. First quadrants from prototypical analysis for all experimental conditions**

	Conventional			Internet		
	Element	<i>f</i>	MEO*	Element	<i>F</i>	MEO*
<b>Football</b>	Goal	107	2.15	Ball	101	1.92
	Ball	97	2.08	Goal	77	2.26
				Club	52	2.88
	Elements = 15 Analyzed evocations = 902 (83.9%) Minimum frequency = 18 Mean element frequency = 60.13 Mean evocation order = 2.9			Elements = 17 Analyzed evocations = 820 (82.2%) Minimum frequency = 17 Mean element frequency = 48.24 Mean evocation order = 2.9		
<b>Aging</b>	Wisdom	136	2.09	Wisdom	113	2.17
	Illness	74	2.49	Illness	78	2.36
	Time	64	2.16	Time	49	2.33
		Elements = 15 Analyzed evocations = 699 (79.7%) Minimum frequency = 19 Mean element frequency = 46.60 Mean evocation order = 2.6			Elements = 16 Analyzed evocations = 660 (79.1%) Minimum frequency = 16 Mean element frequency = 41.25 Mean evocation order = 2.6	

\* MEO = Mean evocation order

Fuente: autores.

to each condition's mean values. Such elements are quantitatively salient and promptly recalled; therefore, the representations' central cores for the investigated samples can probably be found among them. More precise identification would require further methods.

Observing results for the two social objects, a greater similarity in the results of both data collection procedures can be perceived for the aging social object. For the football object there is a coincidence of two elements: goal and ball, while participants that answered through the internet also presented club on the first quadrant. Also, ball and goal have inverted their rank concerning element frequency. For the aging social object, the first quadrants of both conditions are composed by exactly the same elements: wisdom, illness and time, and their frequency ranks have the same order.

### Discussion

The first observed consequence of internet research, in the terms with which it was carried out in the present study, is related to a decrease in coverage precision concerning the desired social group. More than 10% of participants did not belong to the target population. In the present research there was an item that made it possible to identify this problem and remove participants from the sample, and this seems to be a minimum measure to maintain results reliability in social representations studies through the internet. If there was not such control there would be a distortion in results and the representational reconstitution would not be trustworthy. With more sophisticated data systems as a background for the internet form and with access to information about university enrollment numbers, it would be advisable to ask respondents for enrollment numbers or academic passwords.

Considering students' results, it can be observed that it is possible to obtain samples with a good diversity level with relative ease. It is a positive characteristic of studies in which there is a desire to avoid the concentration of participants with certain characteristics.

It was verified as well that internet-based data collection, with online community recruitment,

generates results that are more or less compatible with those emerging from conventional procedures according to the social object's nature. It can be observed that for the aging social object there was greater similarity between results obtained by means of the two data collection procedures, both for sample characteristics and for the representation reconstitution, in spite of the fact that prototypical analysis is an exploratory and somewhat "soft" technique, if considered at an experimental setting.

For the football social object, more disagreeing results were obtained: on the internet more men participated, participants were older, and there was a different distribution on the proximity with football. As for the social representation results for the internet sample indicated a new element on the internet sample: club.

How to explain those results? Seemingly, football is a research social object that involves self-selection more decisively, since participants with different characteristics got interested in and decided to take part on the study, which brought about different results. If conventional methods are to be taken as a parameter for the description of the 'real' representation, self-selection effect must be considered as a problem. Perhaps such is not the case, and to a variety of projects a research carried out through the internet could cover the population of interest more adequately. By allowing anonymity and a different interaction context, internet-based research might favor the expression of what participants really think about a given issue. Nevertheless, considering that conventional paper and pencil techniques are well established in methodological terms, obtaining close results for both response conditions seems to be important in the cases in which there should be no reason to expect differences.

To explain the influence of self-selection for the football social object, it is possible to present the hypothesis that it is a more affective issue than aging for the study's population of interest. The inclusion of the club element to the representation on football's first quadrant, for the internet questionnaire group, corroborates that view: participants that are more interested in football or involved with the sport usually support a club, and their team is

an important element on their representation, for it is through the club supporter group membership that they live in contact with the world of football. According to Reips (2002), self-selection effect is stronger when questionnaires or experiments carried out in the internet deal with polemic or controversial themes. Things do not seem to work differently in social representations research.

Due to the differences both in participants characteristics and representations structure—or more precisely, a first approximation of its structure—, we are led to believe that the internet and conventional questionnaires on football have not covered exactly the same population. There are people that like or dislike football more clearly, while aging, in principle, is not something that generates immediate rejection or approval. An attitudinal component would interfere in the motivation to participate; individuals with an explicit position on a research theme would like to make themselves heard, while less-moving themes would not cause this process, and due to that the results related to them would be closer to those obtained in ‘normal’ conditions, in which the decision about participating or not participating of a questionnaire study is more restricted.

As a result, there is at least one practical derivation of the self-selection effect in socio-representational studies. In order to ensure higher reliability on results validity, the researcher needs to be able to evaluate the degree with which the social object of interest is affectively loaded, or its potential to be affectively moving to the target group or at least to others that might have access to the form. If that is not possible through a bibliographical review of other studies with the populations of interest, the inclusion of attitude scales or items aiming at measuring participants’ positioning and its polarization

is recommended to give a context to results. Internet-based research presents itself as a safer option when used for studies about more neutral objects.

Another option that is available to social representation researchers is to complement in-person data collection procedures with internet-based ones. It is then possible to verify the stability of results in different interaction contexts. It is, in other words, a methodological triangulation resource (Apostolidis, 2003), which allows for wider coverage of representational levels.

It is concluded that internet-based data collection is useful to diversify sample characteristics, but it is essential to adopt measures to restrict participation only to the target population. Results point out that it is even more necessary to take into account the social object’s nature when conducting research through the Web, to avoid exacerbated self-selection effects due to motivation, which could bias results.

It must be emphasized that the present study evaluated only one variety of internet-based research, by means of open-ended questionnaires and discussion forum recruitment. There is a large possibility of research designs, developed to solve or reduce specific survey and experiment data collection problems in internet environments (see, for example, Birnbaum, 2000).

Still, the ease of access to some specific social groups and the possibility of achieving large samples in relatively short time, among other advantages, are very positive aspects of internet-based data collection. It is a technique that might present itself very useful scientifically, as its importance is already growing for other sectors within social psychology and psychology.

## References

- Abric, J.-C. (1998). A abordagem estrutural das representações sociais. In A. S. P. Moreira & D. C. de Oliveira (ed.), *Estudos interdisciplinares de representação social* (pp. 27-38). Goiânia: AB.
- Andrade, A. L. & Wachelke, J. F. R. (2006). Pesquisa psicológica via internet: satisfação em relacionamento de casal e bem-estar psicológico, validação e parâmetros psicométricos. In *Anais da 58ª Reunião Anual da SBPC*. Florianópolis: Sociedade Brasileira para o Progresso da Ciência.
- Apostolidis, T. (2003). Représentations sociales et triangulation: enjeux théorico-méthodologiques. In J.-C. Abric (ed.), *Méthodes d'étude des représentations sociales* (pp. 13-35). Ramonville Saint-Agne: Érès.
- Best, S. J., Krueger, B., Hubbard, C. & Smith, A. (2001). An assessment of the generalizability of internet surveys. *Social Science Computer Review*, 19(2), 131-145.
- Birnbaum, M. H. (ed.) (2000). *Psychological experiments on the internet*. San Diego: Academic Press.
- Couper, M. (2000). Web surveys: a review of issues and approaches. *Public Opinion Quarterly*, 64(4), 464-494.
- Fricker, S., Galesic, M., Tourangeau, R. & Yan, T. (2005). An experimental comparison of web and telephone surveys. *Public Opinion Quarterly*, 69(3), 370-392.
- Hayslett, M. M. & Wildemuth, B. M. (2004). Pixels or pencils? The relative effectiveness of web-based versus paper surveys. *Library & Information Science Research*, 26, 73-93.
- Instituto Brasileiro de Geografia e Estatística (2006). *Pesquisa Nacional por Amostra de Domicílios 2005*. Brasília: IBGE. (Access on 17/04/2006, from IBGE –Instituto Brasileiro de Geografia e Estatística–): <<http://www.ibge.gov.br>>.
- Jodelet, D. (1989). Les représentations sociales: un domaine en expansion. In D. Jodelet (ed.), *Les représentations sociales* (pp. 31-61). Paris: PUF.
- Kaplowitz, M. D., Hadlock, T. D. & Levine, R. (2004). A comparison of web and mail survey response rates. *Public Opinion Quarterly*, 68(1), 94-101.
- Krantz, J. H. & Dalal, R. (2000). Validity of web-based psychological research. In M. H. Birnbaum (ed.), *Psychological experiments on the internet* (pp. 35-60). San Diego: Academic Press.
- McGraw, K. O., Tew, M. D. & Williams, J. E. (2000). The integrity of web-delivered experiments: can you trust the data? *Psychological Science*, 11(6), 502-506.
- Moscovici, S. (1961/1976). *La psychanalyse, son image et son public*. Paris: PUF.
- Musch, J. & Reips, U.-D. (2000). A brief history of web experimenting. In M. H. Birnbaum (ed.), *Psychological experiments on the internet* (pp. 61-87). San Diego: Academic Press.
- Reips, U.-D. (2000). The web experiment method: advantages, disadvantages, and solutions. In M. H. Birnbaum (ed.), *Psychological experiments on the internet* (pp. 89-117). San Diego: Academic Press.
- Reips, U.-D. (2002). Theory and techniques of Web experimenting. In B. Batinik, U.-D. Reips & M. Bosnjak (eds.), *Online social sciences* (pp. 229-250). Seattle: Hogrefe & Huber.
- Vergès, P. (1992). L'évocation de l'argent: une méthode pour la définition du noyau central de la représentation. *Bulletin de Psychologie*, 45(405), 203-209.
- Wachelke, J. F. R. & Andrade, A. L. (2006). Coleta de dados de pesquisa pela internet: validação de duas escalas psicométricas sobre identificação de torcedores de futebol. In *Anais da 58ª Reunião Anual da SBPC*. Florianópolis: Sociedade Brasileira para o Progresso da Ciência.
- Wagner, W. (1998). Sócio-gênese e características das representações sociais. In A. S. P. Moreira & D. C. de Oliveira (eds.), *Estudos interdisciplinares de representação social* (pp. 3-25). Goiânia: AB.

**Fecha de recepción: 26 de febrero de 2009**  
**Fecha de aceptación: 8 de septiembre de 2010**