Increasing participation rates and completeness of questionnaire compilation in web survey. An experimental design of research

Alessandra Decataldo, Department of Sociology and Social research, University of Milan Bicocca, e-mail: alessandra.decataldo@unimib.it

Federico Denti, Department of Sociology and Social research, University of Milan Bicocca, e-mail: federico.denti@unimib.it

Keywords: experimental design; reminder; data linkage; response rate; measurement error

PLEASE, DO NOT QUOTE WITHOUT AUTHORS' PERMISSION

Introduction

Nowadays, there is a widespread use of Internet devices in the field of social research. Consequently, market research companies, national statistical institutes and academic researchers have almost completely abandoned the interviewer-administered surveys. They are confronted with complicated choices regarding the web survey design, since they want to ensure high participation rates and high quality data. Indeed, Lozar Manfreda et al. (2008), based on a meta-analysis of 45 studies which are examining differences in the response rate among web surveys and other survey modes, affirm that the average response rate in web survey is approximately 11% lower than the rate of other survey modes. Instead, the literature that analyses the strategies for the improvement of the data quality in web surveys, still seems to be insufficient (Lynn and Kaminska 2013; Mavletova 2013; Lugtig and Toepoel 2015).

Survey methodologists adopt various strategies to improve participation rate and data quality in general, such as e-mail and SMS (Short Text Message) in forms of pre-notifications, invitations and reminders. Indeed, the declining participation rates and inadequate data quality are serious problems in the field of Internet research, where the absence of interviewers may

pose an additional threat to interviewees' propensity to respond and to complete correctly a questionnaire.

In this paper we use email invitation for all the subjects, but we focus on the impact of: 1) e-mails and SMS reminders on survey participation and data quality; 2) data linkage on survey participation and data quality.

Our web survey aims at students registered during the AY 2015/16 in Bachelor courses at the University of Milan Bicocca (young and educated persons, who are comfortable with the use of Internet).

Within the survey research is widely acknowledged the role of the mobile phones and their functionalities. Wide attention has been paid to the association between the response rates and the data quality of the questionnaires that were filled out by using various devices (Toepoel, Das and Van Soest 2009; Guidry 2012; de Bruijne and Wijnant 2013; Lynn and Kaminska 2013; Mavletova 2013; Mavletova and Couper 2013; Wells, Bailey and Link 2014; Lugtig and Toepoel 2015).

SMS messages certainly represent a new potential provided by mobile phones for the social research. Above all, SMS are used with an aim to improve survey participation. Numerous studies are focusing their attention on this potential by using the SMS as pre-notifications, invitations and reminders (Steeh et al. 2007; Virtanen, Sirkiä and Jokiranta 2007; Bosnjak et al. 2008; Maxl et al. 2010; Fuchs 2012; Mavletova and Couper 2014; Tolonen, Aistrich and Borodulin 2014; Sala, Respi and Decataldo 2014; 2015) and comparing them with other instruments, in particular with the e-mails. However, the results of such types of research are frequently contradictory; better results are sometimes achieved with the use of SMS, sometimes with the use of e-mails, and sometimes there is no difference.

Yet, there seems not to be a single study which lingers over the possibility that SMS in function of pre-notifications, invitations and reminders could improve the quality of questionnaire completion.

Data linkage is a procedure which allows the joining of administrative data with the data derived from the sample surveys. It offers undoubted advantages: limited questionnaire length and reduction of the effort required for the question completion, contributing in that way to the positive consequences concerning the data quality and the reduction of discouragement in terms of refusal and breaking off. We have not found surveys on the possible positive effects of the data linkage with reference to the response rates and data quality.

Our research is a complex experiment, assessing the effects of different combinations of reminder and data linkage on participation and accuracy of filling out the questionnaire (in particular on measurement error).

1.   Research aims

The paradigm of Total Survey Error (Biemer 2010; Groves and Lyberg 2010) provides a general model for the conceptualisation, analysis and assessment of total error of the sample survey.

Concerning the errors within the sample surveys, there is usually made a distinction among sampling errors, the cause of which is known, and non-sampling errors. Non-sampling errors include the specification, sample list, non-response, measurement, coding errors. Errors that are of great relevance for our scopes are non-response and measurement error.

Measurement error is composed of various elements that might individually as well as jointly cause the errors. Indeed, this type of error might be related to the way in which the

questionnaire is administered, to the questionnaire itself, to the context within which the questionnaire is administered, to the interviewer (if he/she is present) and to the interviewed. In particular, planning of questionnaire is a product of a careful reflection by the survey methodologists, since minor editorial changes or changes in the sequence of questions and/or response categories, can affect the responses provided by the interviewed, causing random and systematic measurement errors.

With reference to the non-response errors, there is a very low probability that all members of the sample will participate in a survey. Under some circumstances, e.g. when the non-response rate is high and the respondents have different characteristics respect to the non-respondents, the non-response error might occur. In case of a lack of the procedures which should reduce the impact of the non-response rate, it will not be possible to apply to the entire population the results of a survey carried out on a particular sample.

In general, data quality (or measurement error) is intended for the data compliance under the logical and methodological conditions required by the research aims. Therefore, the dimensions of quality concept refer to various aspects: validity, reliability, completeness, congruence and relevance (Alvin 2007; Fricker and Tourangeau 2010).

Validity refers to the relation between concept and indicator (Bryman 2004) and to the selection of indicators. The terms *validity* and *reliability* are frequently considered interchangeable, nevertheless the fact that they refer to different moments within the social research. As a matter of fact, the term *reliability* focuses on the operational definition, data gathering, and information processing. Both of the terms mostly deal with specification error, i.e. with the questionnaire formulation (which is currently not focus of our interest, as it will be specified further in the text).

Completeness refers to three different aspects: completeness of the achieved sample respect to the original one; participation of the respondents throughout all the phases predicted by a research design; respondents' propensity to answer all the questions within the questionnaire.

Relevance deals with the effective pertinence and the importance of the gathered information respect to the research aim and to the specific context within which it is carried out.

In experimental designs the congruence may assume several meanings and its importance can also vary according to the research aims. Usually, there are proprieties which are unlikely to change during the research, such as subjects' socio-demographic proprieties.

Validity and reliability do not need a further reflection, because our research is not related to the questionnaire formulation. Furthermore, within our questionnaire there cannot be identified questions that are of greater relevance respect to the others (our aim was to have a concise questionnaire, which is composed of few identically relevant questions). Therefore, this work focuses on completeness and congruence.

With reference to another issue, i.e. to the response rate, there are many references in literature on this topic, and the debate throughout the recent years has been almost exclusively focused on the surveys carried out via IT devices (computers and smartphones), and, therefore, in form of the web surveys. Furthermore, the aim of most of the articles is to identify the strategies that would enhance response rates (Tourangeau, Conrad, and Couper 2013; Callegaro, Manfreda and Vehovar 2015). AAPOR (2016) defines *response rates* as the number of complete interviews with the reporting units that are divided by the number of eligible reporting units in the sample.

Our aim is to contribute to expansion of the current knowledge concerning the effectiveness of some of the "tools" that methodologists can use to enhance survey participation and to reduce measurement errors. In particular, our hypothesis is that, in order to reach these objectives, two strategies can be used: the use of reminders (which would press the subjects

for an answer and make them feel more motivated to respond in a more complete and, possibly, in a more correct manner), and the reduction of the questionnaire length. The first strategy will be implemented through the use of e-mail and SMS, supposing that SMS is more efficient than e-mail. The second strategy will be implemented through the use of data linkage, in order to avoid the request for the administrative information, which are already available within the university database.

Our focus of interest is the comparison of SMS and e-mail reminders, since these are the contact modes that are often used in Web surveys (Keusch 2015; Millar and Dillman 2011).

In the context of survey methodology, the social exchange theory has been used to explain respondents' participation (Dillman 1978; Dillman Smyth and Melani Christian 2009; Sala, Respi and Decataldo 2014; 2015). According to this theory, respondents' participation in surveys is driven by the need to maximise the reward/cost relationship; the rewards are associated to the fact that their responses are important for the researchers, i.e. the idea that social credit is preferable to social indebtedness. Therefore, we state that respondents may perceive SMS as a more expensive (both socially and economically) contact mode than e-mails, while the combination of SMS and e-mails could be perceived as the most expensive. Moreover, in our specific case, one SMS that is sent to a private number might be perceived as a more personalised and more direct contact mode respect to an e-mail sent to the institutional e-mail address, that is provided by the University. As a sign of appreciation for the researchers' efforts, respondents who received SMS reminder and, even more, who received both kinds of reminder may be more likely to participate and provide more accurate answers.

As it was stated above, it is a well-known fact that sending reminders has a positive impact on the response rate, but we are not familiar with any type of studies that assessed the effects of contact modes on data quality.

6

Concerning the data linkage, it exemplifies an opportunity to improve the completeness of the questionnaire compilation (by reducing the number of items), and to reduce the percentage of inconsistent data (i.e. the percentage of discrepancy between the information contained within the administrative database and the responses provided by a interviewed inside the questionnaire). As a matter of fact, the procedure of data linkage has the advantage of making more sustainable the compilation for an interviewee.

Even though Mavletova (2013) found no effect of the questionnaire length on completion and break off rates, we think that the reduction of the questionnaire length could have positive effects both on the response rate and on the data quality. There are no researches known to us that have investigated this matter.

2. Data

We have a plain example of *mail survey of specifically named persons* (AAPOR 2016). Reference population is represented by the students enrolled in 3-year-courses at University of Milan-Bicocca in the AY 2015-16.

The first extraction of data regarded 6,179 enrolled students. Since the study aimed to gather the information regarding the students' evaluation on the basis of the first impact that the university reality had on them, there were chosen only 5,389 individuals that had enrolled for the first time in AY 2015-16 and entered for the first time the university system (therefore, no previous experience related to university courses, even if they hadn't been completed and the degree hadn't been obtained). Among them there were eliminated 116 cases which were not included within the dataset provided by the Statistical Office of the University, containing some information which were considered sensitive and necessary for the realisation of the

study (these are the people that might have performed the pre-enrolment at university, without having it concluded afterwards with a real and proper enrolment).

From the moment in which the research aimed to study also the effectiveness of the reminders received via an SMS, the remaining 4,927 cases were forwarded to the system of SMS sending management (www.smshosting.it), which identified 112 invalid telephone numbers[1]. The remaining 4,815 cases represent the reference population for this study.

It is known that the surveys have to establish a useful data which should determine the eligibility/adequacy of cases. This usually coincides with the first day of the field research. In our specific case, since this is a research specifically oriented to gather the information related to the first year course performance, there were considered as eligible subjects also those students that decided to drop out from the university. We know the precise number of these students (88, which corresponds to 1.8% of the reference population), because it is contained within the administrative information that the University put at our disposal. Therefore, the structure of our population implies that there are no cases of the unknown eligibility.

3. Methodology

In order to check our hypothesis, we consider eight experimental groups:

- T1.1-SMS, that will receive SMS only as a reminder and a questionnaire with additional questions concerning the administrative information;

---

[1] With reference to the e-mail addresses, those are all valid since they are institutional e-mail addresses assigned by the University at the moment of enrolment.

- T1.2-SMS, that will receive SMS only as a reminder and a questionnaire without additional questions concerning the administrative information (through the use of the data linkage only);

- T2.1-E-mail, that will receive an e-mail only as a reminder and a questionnaire with additional questions concerning the administrative information;

- T2.2-E-mail, that will receive an e-mail only as a reminder and a questionnaire without additional questions concerning the administrative information (through the use of the data linkage only);

- T3.1-E-mail and SMS, that will receive both SMS and an e-mail as a reminder and a questionnaire with additional questions concerning the administrative information;

- T3.2-E-mail and SMS, that will receive both SMS and an e-mail as a reminder and a questionnaire without additional questions concerning the administrative information (through the use of the data linkage only);

- C.1-no reminder, that is the control group and it will not receive any reminders, but a questionnaire with additional questions concerning the administrative information;

- C.2-no reminder, that is the control group and it will not receive any reminders, but a questionnaire without additional questions concerning the administrative information (through the use of the data linkage only).

Allocation of 4,815 cases into eight experimental groups was achieved with the use of a random criterion. The first group was composed of 601 cases, and each of the other seven groups was composed of 602 cases.

After the allocation of cases into eight groups, the composition of the groups was verified on the basis of the known socio-demographic characteristics of the subjects that were part of them (administrative information provided by the University of Milan-Bicocca). From the moment in which it is plausible to hypothesise that the presence or absence of such

characteristics could affect the propensity to participate in the research, and it is usually difficult to separate measurement effects from selection effects (Shadish, Campbell and Cook 2002), it is necessary to verify that the groups' structure, which is the result of a random allocation, is homogeneous respect to such characteristics.

The variable that defines the experimental group membership was related to the following variables, deriving from the administrative sources and, therefore, belonging to the entire reference population: sex, age at enrolment (aggregated in three modalities: 18-19 years old, 20-21 years old, 22 years old or over), citizenship (Italian, non-Italian), residence area (province of Milan, other provinces of Lombardy, other regions), high school graduation (2015, 2014, 2013 or before), type of high school graduation (classical high school, scientific high school, other high schools or teachers college, technical institute, vocational institute, foreign school), exit examination mark (in cents: 60-70, 71-80, 81-90, 91-100), student's status at the moment of data extraction (enrolled, dropout), school/department where a student is enrolled (Economics and Statistics, Law, Medicine and Surgery, Psychology, Science, Education Sciences, Sociology). The analyses (Tab. 1) have shown the absence of particular over-representations of some population segments within the experimental groups, confirming in that way the substantial equivalence of the groups themselves[2].

In order to explore our hypothesis, we used two versions of the questionnaire; one that can be defined as a short questionnaire and the other that is called a long one. In the first version, by exploiting the potential of the data linkage, we did not ask questions related to the information that are already available within the Bicocca dataset (those data that are reported in tab. 1); this version is composed of 13 questions. In the long version, there were introduced questions

---

[2] Contingency tables, that relate the variable which defines the experimental group to the 9 socio-demographic variables, show only small percentage differences and the chi-squared value (reported in the last column of the table 1), which is insignificant as a new proof of absence of association between the experimental group membership and each of the variables.

related to the information that are already available within the Bicocca dataset; this version is composed of 20 questions, i.e. it has 7 extra questions added at the end of questionnaire.

The questionnaire is composed of 20 questions, 11 of which are referring to the subject matter of this research and 9 of them are socio-demographic. 11 content questions consist of 7 question sets, 3 single-answer questions, 1 multiple-choice question. There is no possibility to answer the questions with "other" or "don't know".

Overall, 79 variables can be developed out of 11 questions. All the individuals were asked only two out of 9 socio-demographic questions (all being single-answer questions); the components of four experimental groups with the long questionnaire were also asked the other 7 questions. Finally, there could be identified 88 variables for the experimental groups T1.1, T2.1, T3.1 and C.1, and 81 variables for the experimental groups T1.2, T2.2, T3.2 and C.2.

Tab. 1 – Composition of the experimental groups according to the socio-demographic characteristics (population, %)

	T1.1	T1.2	T2.1	T2.2	T3.1	T3.2	C.1	C.2	Total	Chi-squared, Significance
Sex										0.781
Male	42.4	42.5	42.7	42.0	43.2	43.9	45.3	40.0	42.8	
Female	57.6	57.5	57.3	58.0	56.8	56.1	54.7	60.0	57.2	
Age at enrolment										0.186
18-19 years old	67.6	67.1	64.8	66.4	68.9	63.6	63.3	65.9	66.0	
20-21 years old	25.3	24.6	25.7	21.8	23.3	25.2	27.2	24.3	24.7	
22 years old or over	7.2	8.3	9.5	11.8	7.8	11.1	9.5	9.8	9.4	
Citizenship										0.167
Italian	93.5	91.7	93.0	95.7	94.2	92.7	93.7	94.5	93.6	
Non-Italian	6.5	8.3	7.0	4.3	5.8	7.3	6.3	5.5	6.4	
Residence										0.729
Province of Milan	33.4	34.6	36.7	36.9	38.0	37.7	34.1	36.5	36.0	
Other provinces of Lombardy	53.6	50.3	50.3	50.7	50.3	50.3	52.7	48.7	50.9	
Other regions	13.0	15.1	13.0	12.5	11.6	12.0	13.3	14.8	13.1	
Year of graduation										0.158
2015	78.2	80.6	79.4	78.4	80.2	77.6	78.6	81.2	79.3	
2014	13.6	11.1	12.1	9.8	12.0	10.6	11.3	9.5	11.3	
2013 or before	8.2	8.3	8.5	11.8	7.8	11.8	10.1	9.3	9.5	
Type of graduation										0.751
Classical High School	5.8	5.8	5.6	5.6	5.1	5.8	3.8	5.8	5.4	
Scientific High School	38.9	33.2	37.9	35.5	35.2	32.1	35.9	35.5	35.5	
Other High Schools – Teachers College	24.6	25.2	21.6	24.9	24.6	23.8	24.9	23.6	24.2	
Technical Institute	23.0	26.1	26.4	26.6	25.2	27.4	26.7	26.2	26.0	
Vocational Institute	5.0	7.1	6.6	5.6	7.8	7.6	6.3	7.5	6.7	
Foreign School	2.7	2.5	1.8	1.7	2.0	3.3	2.3	1.3	2.2	
Exit exam mark										0.351
60 – 70	33.5	33.4	31.7	33.1	31.6	31.9	37.9	30.7	33.0	
71 – 80	29.8	33.6	35.2	33.8	33.6	32.9	29.7	33.4	32.8	
81 – 90	22.8	21.6	20.9	20.5	24.9	21.5	21.1	24.8	22.3	
91 – 100	13.8	11.5	12.1	12.6	10.0	13.6	11.3	11.1	12.0	
Status at the moment of data extraction										0.507
Enrolled	98.7	97.8	98.3	98.2	98.8	97.2	98.3	98.0	98.2	
Dropout	1.3	2.2	1.7	1.8	1.2	2.8	1.7	2.0	1.8	
School/Department										0.289
Economics and Statistics	24.1	23.9	21.9	23.3	24.9	26.9	23.3	21.4	23.7	
Law	2.7	3.5	3.7	3.2	4.3	3.3	4.0	3.0	3.4	
Medicine and Surgery	8.8	8.5	10.0	11.1	7.8	8.8	8.0	8.0	8.9	
Psychology	10.0	12.1	8.3	9.3	10.3	9.6	10.0	10.8	10.1	
Science	29.5	23.8	27.6	26.2	28.1	26.1	28.6	27.7	27.2	
Education Sciences	14.3	13.3	16.6	13.6	10.8	15.0	12.5	13.3	13.7	
Sociology	10.6	15.0	12.0	13.3	13.8	10.3	13.8	15.8	13.1	
N	601	602	602	602	602	602	602	602	602	

The first request for the participation to the web survey was sent to the institutional e-mails of the enrolled students on Thursday, 7th July 2016, at 11 o'clock. A reminder, the type of which depended on subjects' group membership, was sent on Thursday, 14th July at 11 o'clock, to all the subjects that had not answered or had filled out the questionnaire only partially. In particular, on the basis of the experimental design, to the components of the groups T1.1-SMS and T1.2-SMS was sent a reminder via SMS; members of the groups T2.1-E-mail and T2.2-E-mail received a reminder via email; members of the groups T3.1-E-mail and SMS, and T3.2-E-mail and SMS received both types of reminders, while for the control groups C1-no reminder and C.2-no reminder was not considered any kind of reminder. Overall, there were sent reminders (via email and/or SMS) to 3,355 individuals. Among them, there were not delivered 77 e-mails (due to the disabled account) and 59 SMS (due to the fact that the cell-phones were switched off or out of reach for more than 24 h)[3]. Data gathering was concluded on Thursday, 21st July, at 11 o'clock.

Bellow follows the text of the e-mail request that was sent with lsa.sociologia@unimib.it as a sender:

Dear {FIRSTNAME} {LASTNAME},

Research Laboratory of Applied Sociology of the University is carrying out a research on the service and facilities fruition of Milan Bicocca by the enrolled students at the end of the first course year. The aim of the research is to suggest the interventions which would contribute to the improvement of the university experience for our students.

---

[3] The 77 e-mails mentioned above were rejected only when in function of reminders, and not when they were sent as a request to participate to a research. It can be hypothesised that during the week under discussion, the University dismissed e-mail accounts of the enrolled students who interrupted their studies in the course of the actual academic year.

The focus of research are all the students that enrolled in the AY 2015-16. We kindly ask you to collaborate and to fill out the on-line questionnaire by clicking on the following link; we ask you to do it even in case you have decided not to proceed with the studies at our University: {SURVEYURL}.

Obviously, we can guarantee that your answers will only be used in the aggregate manner and in full compliance with data protection laws regulating the processing of personal data (Legislative Decree no. 196 from 30/6/2003).

We thank you in advance for your collaboration.

Yours sincerely,

Prof. Alessandra Decataldo (Person in charge of the research)

Dr. Federico Denti (manager of the Applied Sociology Laboratory)

For more information contact:

lsa.sociologia@unimib.it

Tel. +390264487518

http://www.sociologiadip.unimib.it/lsa

The reminder sent via e-mail was sent by the same sender and it contained the text that follows (we did not include the link to the questionnaire within the e-mail in order to maintain the equal stimulus provided by two different types of reminders):

Dear {FIRSTNAME} {LASTNAME},

We inform you that there is only one week left to fill out the questionnaire by clicking on the link that was sent in the previous e-mail, received on 7th July.

We remind you that the research focuses on the service and facilities fruition of Milano Bicocca by the enrolled students, and its aim is to suggest the interventions which would contribute to the improvement of the university experience for our students. Participation of all the subjects to whom the request was sent is of fundamental importance for the successful completion of our research.

We thank you in advance for your collaboration.

Yours sincerely,

Prof. Alessandra Decataldo (Person in charge of the research)

Dr. Federico Denti (manager of the Applied Sociology Laboratory)

For more information contact:

lsa.sociologia@unimib.it

Tel. +390264487518

http://www.sociologiadip.unimib.it/lsa

Within the reminder (obviously a shorter one, due to a limited number of characters that can be used within one SMS) sent via SMS the sender was Uni-Bicocca (short name for University of Milan Bicocca), and it contained the following text:

You have one week left to fill out the questionnaire on university experience by the enrolled students: use the link from the e-mail received on 7/7. THANK YOU!

3.1 Assessing response rates

According to AAPOR (2016), we define:

RR= Response rate

I = Complete interview

P = Partial interview

R = Refusal and break-off

NC = No-contact

As it has already been explicitly stated above, in the paragraph Data, the following cases do not make part of our data:

O = Other

UH = Unknown if household/occupied

UO = Unknown, other

e = Estimated proportion of cases with the unknown eligibility that are eligible

In order to explain what is meant by the expression *complete interview*, we always refer to AAPOR (2016). In particular, less than 50% of all the answered questions (with other than refusal or no answer) refer to break-off, 50%-80% are partial and more that 80% are complete. The questionnaire was composed only of questions that could have been answered by the interviewed; there were no filter questions.

The NC are present only with reference to the reminders; they are those subjects to whom a reminder e-mail was not delivered or whose telephone number resulted to be out of reach.

The so-called O do not exist in our case, since (except for the case of death) it is highly improbable that a student has difficulties indicated by AAPOR (2016), or suffers from a mental disease or has linguistic difficulties that prevent him/her from answering the questions. Therefore, the response rates useful for our scopes are (AAPOR, 2016):

$$RR5 = \frac{I}{(I + P) + (R + NC + O)}$$

$$RR6 = \frac{I + P}{(I + P) + (R + NC + O)}$$

3.2 Assessing data quality

Most of the studies use the indirect methods in order to assess measurement error in a survey research. Indirect methods link measurement error in surveys to the process of answering a survey question (Ludwig and Toepoel 2015). For example, if a respondent uses frequently the "Don't know" answer category, it is highly probable that this respondent does not put a lot of cognitive effort in the performance of the question-answer process (Tourangeau, Rips and Rasinski 2000).

Measures that were used in literature to detect measurement errors in web surveys are (Lynn and Kaminska 2012; Mavletova and Couper 2013; Ludwig and Toepoel 2015; Wells, Bailey and Link 2014):

- "do not knows": a greater presence of "do not know" answers implies a greater use of cognitive shortcuts among the respondents, and the presence of larger measurement errors.

- Length of open answers: shorter (and less substantive) answers stimulate the respondents to satisfice, and, as a result, measurement errors are larger.

- Consistent answer patterns (straightlining): if the respondents keep responding in a consistent way to every item of a question set, it is more likely that they will satisfice. It can be seen in a form of respondents agreeing to every Likert-scale answer option, or of the respondents choosing a consistent "extreme" or "middle" response category.

- Primacy effect: choosing the first answer indicates satisficing and increased measurement error.

- Rounding: there is a greater probability that the respondents will satisfice if they keep rounding continuous answers to the nearest whole number.

- Answers to sensitive items: lower reports of sensitive behavior and attitudes indicate satisficing in web survey.

According to the literature on measurement error, Lugtig and Toepoel (2015) advise to use as indicators: item missing, open questions (short or no answers are a proxy for more

measurement error), straightlining (response set), primacy effect, number of answers checked in a check-all-that-apply question, forms of paradata (interview duration and respondents' evaluation of the questionnaire).

As we have specified at the beginning of this paper, our research aims to focus on the measures of data completeness and congruence. We also lack validation data, which means that only indirect indicators of measurement error can be used. For that scope we identified a series of indicators, relying on those that have just been cited as those that are used the most in literature. Due to the way in which we decided to structure the questionnaire, such indicators did not allow the "don't know" answers (within the questions, there was voluntarily not inserted the "don't know" answer modality), length of open answers (open questions were not included on purpose), primacy effect (LimeSurvey does not permit us to know the order of the administrated questions), and answers to sensitive items (there were no sensitive questions).

Regarding the **data completeness**, there are several aspects to focus on, therefore five indicators were calculated to monitor this issue:

- **A1 - % missing** – is calculated as a percentage ratio between the number of missing answers and the number of questions. Sets of questions are considered as single questions, and the answers are defined as missing only when there is no answer for all the items of a question set. For the experimental groups with a long questionnaire (T1.1, T2.1, T3.1, C.1) the number of the considered variables is 20, whereas for the groups with a short questionnaire (all the others that are exploiting the data linkage) the number of variables is 13.

- **A2 - % partial missing** – is calculated as a percentage ratio between the number of the missing answers to the question sets and the overall number of items within the question sets. There have been excluded questions that differ from the question sets.

Number of the considered variables is 70 for all the respondents because there are no question sets among the additional questions for the experimental groups with a long questionnaire.

- **A3 - % item no response** – is calculated as a percentage ratio between the number of missing answers and the overall number of variables. Therefore, this indicator comprises single-answer questions, all the items of question sets, and the multiple-choice question. For the experimental groups with a long questionnaire, the number of the considered variables is 88, while for the groups with a short questionnaire it is 81.

- **A4 - % response set** – is calculated as the percentage ratio between the number of homogeneous answers given to the items of each question set and the overall number of the items of the question sets. From the moment in which there are no question sets among the additional questions for the members of the experimental groups with a long questionnaire, the number of the considered variables is 70 for all of them.

- **A5 - % personal data missing** – is calculated as a percentage ratio between the number of the missing answers and the number of questions which are limited to socio-demographic variables. The indicator can be calculated only for the members of the experimental groups with a long questionnaire, who were considered for such questions. The number of the considered variables is 9.

The other critical issue regards the **congruence** of the gathered information, which is intended as data coherence between administrative information and gathered data. It is applicable only to the respondents with a long questionnaire, because it gives us the possibility to confront the congruence of the administrative information, that were declared at the moment of the enrolment, with those declared to us during the interview. The congruence was calculated through the amount of the inconsistent data:

-   **C1 – inconsistent data %** - is calculated as the percentage ratio between the detected discrepancies within the respondents' answers and the administrative data for the total of the socio-demographic questions. A missing answer is also considered to be incoherent. The indicator can be calculated only for the members of the experimental groups with a long questionnaire, who were considered for such questions. The number of variables is 7, and it corresponds to those that are available for the administrative data.

A further control of the data quality that was accomplished is the calculation of the interview duration:

-   **duration** – Time for filling out (in minutes). It corresponds to the difference between the date of the first questionnaire completion (day and hour) and the date of the last questionnaire completion, even in cases when the filling out was incomplete. The values over 30' (a duration that is considered to be a threshold of plausibility), which are supposedly a result of the interruptions during the questionnaire completion, were reported as a value of 30'. As a matter of fact, the durations which exceeded 30' seemed to be more compatible with the respondents leaving the questionnaire open for a long time, without answering any questions or answering the questionnaire in several phases.

In order to reach our aims, we mainly perform bivariate analysis and compute appropriate statistical tests to check statistically significant differences among groups (Pearson's Chi Square Test for Independence and ANOVA F-Test). All the analyses were conducted on the complete and partial interviews, i.e. on the questionnaires with 50% of the answered questions.

4. Results

4.1 Response rates

Coherently to what has been indicated in the previous paragraph, for the calculation of our response rates we used the following data:

I = Complete interview = 810

P = Partial interview = 86

R = Refusal and break-off = 3,919

NC = No-contact (only for the reminder) = 135

Therefore, we obtained the following general response rates (for the sample of the respondents at the completion of information gathering):

RR5 Response rate for the complete interviews = 0.168 = 16.8%

RR6 Response rate for the complete interviews + partial interviews = 0.186 = 18.6%

The obtained response rates are low, but this is understandable if we take into consideration a restricted time window (only 15 days) in the course of which the web survey was active, as well as the fact that data gathering occurred in July, the month during which in Italy there are no university lectures and the summer holidays begin. However, low response rates do not represent a problem from the moment in which our research has no aspirations to generalise the substantive results (with regard to the fruition and evaluation of the first year of studies), but it aims to evaluate the difference between different experimental groups with regard to response rates and measurement errors.

Table 2 reports the response rates RR6 (the maximum response rate) which are calculated for every experimental group at the end of the first week (before the reminder) and at the end of the second week (after the reminder). As it can be noticed, after the first week the response rates are very similar among different experimental groups (the maximum difference is 3 percentage points) and, above all, there is an insignificant relation between the response rate and the experimental groups membership (measured through the Chi-square). On the contrary, the response rates at the end of the second week (calculated according to the net number of the respondents from the first week and the number of those that were not contacted) show evident differences.

Tab. 2 – Response rates before and after the sending of the reminder, for the experimental group

	Response rate before the reminder (a)		Response rate after the reminder (b)	
	%	N	%	N
T 1.1 - SMS with additional questions	7.5	601	20.0	544
T 1.2 - SMS without additional questions	6.1	602	20.9	549
T 2.1 - E-mail with additional questions	7.8	602	4.6	541
T 2.2 - E-mail without additional questions	7.1	602	4.5	535
T 3.1 - SMS and e-mail with additional questions	9.3	602	22.8	526
T 3.2 - SMS and e-mail without additional questions	8.1	602	24.8	525
C 1 - No reminders with additional questions	6.3	602	0.7	568
C 2 - No reminders without additional questions	8.3	602	0.7	559
Total	7.6	4,815	12.2	4,347
Chi-square, significance	0.468		0.000	

Notes: (a) corresponds to the percentage ratio of the number of the respondents that filled out the questionnaire (even partially) before the sending of the reminder and the total number of population; (b) corresponds to the percentage ratio of the number of the respondents that filled out the questionnaire (even partially) in the period after the reminder was sent and the total number of population, to the net number of the respondents that filled out completely the questionnaire before the sending of the reminder (333 cases), and to the net number of the respondents that did not receive the reminder via e-mail and/or SMS (135 cases).

In particular, according to our initial hypothesis, the groups that received SMS have remarkably higher response rates respect to those that received an e-mail (16 percentage

points of difference). The groups that received double reminders, coherently with our hypothesis, show an even more remarkable advantage. The groups without reminder keep the response rates substantially unchanged after the date in which the reminders were sent to the members of the other groups.

Furthermore, in the four experimental groups in which the reminder proved to be of a greater efficiency, as we have hypothesised, the data linkage (by shortening the length of the questionnaire) represents a further incentive to fill out the questionnaire. In the groups that received SMS as a reminder, the percentage of the difference is only 0.9%, while in the groups that received both SMS and an e-mail, 2 percentage points refer to the difference.

Finally, there is a statistically significant association between the response rate at the end of the second week and the experimental group membership.

4.2 Measurement error

The results related to the indirect indicators of the measurement error show a good overall quality of the questionnaire completion. Except for the duration of filling out (which is an average of the duration of questionnaire filling out), the values of the indicators represent percentage ratios which were developed according to what has been indicated in Paragraph 3.2.

Indeed, with reference to the data completeness:

- **A1 - % missing**: 63.4% of the respondents (total of 896 respondents) did not miss a single question and the overall average of the sample is only 5.7% of the questions that were completely missed.

- **A2 - % partial missing**: 23.3% of the respondents provided answers for all the items of the question sets and the average is 7.9% for the missing questions in question sets, respect to the overall number of the items within the question sets.

- **A3 - % items no response**: 19.9% of our respondents did not miss a single answer and the average of the overall sample is only 7.4% missing answers respect to the overall total of the variables.

- **A4 - % response set**: 81.7% of the respondents have no response set and the overall average of the sample is only 4.9% of the response set for all the question sets.

- **A5 - % personal data missing**: 83.3% of the interviewed belonging to the groups with a long questionnaire (the only ones who were asked the questions related to personal data, corresponding to 444 cases) do not show personal data missing and the total average of this subsample is 8.3% for the personal data missing.

With reference to the **congruence** between administrative information and gathered data (applicable only to the respondents with long questionnaire):

- **C1 – inconsistent data %**: 80% of the cases provided answers referring to the socio-demographic data which are coherent with the administrative data provided by University of Milan-Bicocca. The average of our subsample is 7.4% for the inconsistent data respect to the total of the questions related to the personal data.

With reference to the further control of the data quality:

- **Duration**: 10.5% of the interviewed spent 30 minutes and more for the filling out, probably due to the fact that they completed the questionnaire in more than one session or they were doing something else and left the questionnaire open. The average time of the filling out for the rest of the sample corresponds to a little less than 10 minutes.

Table 3 shows the results of the ANOVA F-test, which was performed in order to analyse the relation between the indicators and experimental group membership. In this case, since our

attention is focused on the effect of different reminders, the analysis is conducted on only 523 cases corresponding to the individuals that completed the questionnaire after the reminder was sent. As it can be noticed, all the relations result to be insignificant (measured by the F test by Snedecor), except for those with the indicators C1 – inconsistent data and duration (even though the relations seem to be weak, as it is shown by the eta squared values).

The last one (duration) shows that for the groups with a long questionnaire, it takes on average few minutes longer to complete the filling out; however, the result seems to be related to the presence of 7 additional questions in their questionnaire. However, there should be noticed that for the respondents who receive an SMS it takes a bit longer to complete the questionnaire, since they seem to show a greater accuracy while filling it out.

Moreover, with reference to the indicator C1 – inconsistent data, coherently with our hypothesis, the results show that the use of SMS can guarantee a greater coherence of the answers respect to the use of e-mail.

However, for both of the indicators, the joint use of both SMS and e-mail does not achieve the hypothesised effect.

Generally, the results show that the use of the reminders and the data linkage in particular (since there is no significant relation between the calculated indicators for all the respondents and the experimental group membership) achieved the hypothesised effects only marginally. Indeed, the use of the reminder via SMS is associated with the levels of accuracy that are slightly higher only respect to some of the indicators that have been adopted, while the reduction of the questionnaire length, which was obtained through the data linkage, apparently did not have any effects on the level of accuracy for the filling out.

Tab. 3 - Indicators of the measurement error, for the experimental group. Analysis of the variance (a)

	A1 – Item missing related to questions	A2 – Item missing related to question sets variable	A3 – Item missing related to variables	A4 – Response set	A5 – Item missing on the socio-demographic variables	C1 – Coherence of the answers with the administrative data	DURATION – Time for filling out	N (b)
T1.1 - SMS with additional questions	5.41	7.58	7.01	2.79	6.12	4.76	13.95	109
T1.2 - SMS without additional questions	4.75	8.67	7.95	4.52			11.16	115
T2.1 - E-mail with additional questions	8.80	6.17	6.68	2.69	14.67	19.43	13.20	25
T2.2 - E-mail without additional questions	8.97	10.00	9.98	7.74			10.67	24
T3.1 - SMS and e-mail with additional questions	6.79	8.44	7.98	4.87	8.89	7.35	13.08	120
T3.2 - SMS and e-mail without additional questions	4.56	6.85	6.36	7.10			11.67	130
Total	5.70	7.88	7.40	4.94	8.27	7.43	12.38	523
ANOVA:								
F	1.77	0.70	0.81	1.74	1.65	5.30	2.26	
Significance of F	0.12	0.63	0.55	0.12	0.19	0.01	0.05	
Eta squared	0.02	0.01	0.01	0.02	0.01	0.04	0.02	

Note: (a) except for the duration of the filling out, the values represent the average of the values assumed by the indicators, which are calculated as the percentage ratio between the number of the situations of inaccuracy (missing answer, homogeneous answer in a question set, incoherent answer with the administrative datum) and the total number of possible answers for each of the indicators that are considered; (b) the analysis was conducted on the total number of filling outs (complete and partial) performed after the reminder was sent; due to the fact that they were not numerous, there were excluded from the analysis 8 cases that corresponded to the members of the two control groups that completed the questionnaire in the second week of data gathering.

5. Summary and conclusions

In this paper, we have investigated response rate and measurement error in web survey. On the one hand, the literature on the strategies that should be applied in order to increase the response rate in web survey grows constantly, on the other the literature concerning the improvement of the data quality is still in its infancy. In this paper we focused on the impact of: 1) e-mails and SMS reminders on survey participation and data quality; 2) data linkage on survey participation and data quality.

With reference to the use of reminders in order to increase the response rate in web survey, there are various types of research with the results that are not always agreed (Steeh et al. 2007; Virtanen, Sirkiä and Jokiranta 2007; Bosnjak et al. 2008; Maxl et al. 2010; Fuchs 2012; Mavletova and Couper 2014; Tolonen, Aistrich and Borodulin 2014; Sala, Respi and Decataldo 2014; 2015), whereas the research seems to be insufficient in the field of measurement error reduction. We are interested in comparison of SMS and e-mail reminders. According to the social exchange theory (Dillman 1978; Dillman Smyth and Melani Christian 2009; Sala, Respi and Decataldo 2014; 2015), we state that respondents may perceive SMS as a more expensive, direct and intimate contact mode than e-mails, while the use of both SMS and e-mail can be perceived as an even greater effort that the researcher has to make. Consequently, we supposed that respondents who received SMS reminder and, even more, those who received both kinds of reminder, may be more likely to participate and provide more accurate answers.

There is a lack of literature concerning the use of the data linkage in order to increase the response rate and to reduce the measurement error. Since it exemplifies an opportunity to improve the completeness of the questionnaire compilation and to reduce the percentage of inconsistent data, we supposed that by making the questionnaire shorter, there should be achieved positive effects on the response rate and on the quality of answers.

Owing to the disentangling of the selection effects from measurement effects, we managed to see the experimental design with 8 randomized groups. In this way we could rule out selection effects, and focus on measurement error and response rate associated with the type of reminder and data linkage. Indeed, our research is a complex experiment which assesses the effects of different combinations of reminder and data linkage on participation and measurement error. Concerning the response rate, we referred to what was said by AAPOR (2016). With reference to the measurement error, starting from the reference literature (Lynn and Kaminska 2012; Mavletova and Couper 2013; Ludwig and Toepoel 2014; Wells, Bailey and Link 2014), we designed 7 indirect indicators for the measurement error.

Our results, with reference to response rates, show that after the first week the response rates are similar among various experimental groups, and, above all, the relation between the response rate and the experimental group membership is insignificant. On the contrary, the response rates at the end of the second week show noticeable and statistically significant differences. Indeed, in accordance with our initial hypothesis, the groups that received SMS have a remarkably higher response rate in comparison to those who received an e-mail. The groups that received a double reminder seem to be even more advantageous.

Furthermore, for the four experimental groups that are positively stimulated by the use of an SMS reminder, the data linkage, as we hypothesised, seems to be a further incentive for the questionnaire completion.

With reference to the indicators of measurement error, all the relations seem to be insignificant, except for those with the indicators C1 – inconsistent data and duration (but with very weak relations). Even in this case, in coherence with our hypothesis, the results show that the use of SMS guarantees a greater accuracy of the answers respect to the use of e-mail.

With reference to the indicators of the measurement error, it has to be noticed that, according to the results, the use of reminders and, in particular, the use of data linkage, achieved only partially the hypothesised results. As it has already been said, the use of the reminder via SMS is associated with the levels of accuracy that are slightly higher only respect to some of the adopted indicators, while the reduction of the questionnaire length, which was realised through the data linkage, apparently did not affect the level of accuracy of the filling out.

We believe that assessing the impact of different contact modes on measurement error (and not only on response rate) is a very promising research field. In the same way, we consider that it is useful to explore even more profoundly the potential offered by the data linkage.

References

AAPOR, 2016, Standard Definitions. Final Dispositions of Case Codes and Outcome Rates for Surveys Revised 2016: RDD Telephone Surveys, In-Person Household Surveys, Mail Surveys of Specifically Named Persons, Internet Surveys of Specifically Named Persons, https://www.aapor.org/AAPOR_Main/media/publications/Standard-Definitions20169theditionfinal.pdf.

Alwin D.F. (2007), *Margins of error - A study of reliability in survey measurement*, Hoboken, New Jersey, Wiley.

Biemer P.B. (2010), "Total Survey Error: Design, Implementation, and Evaluation", *Public Opinion Quarterly*, 74 (5), 817–848.

Bosnjak M., Neubarth W., Couper M.P., Bandilla W., Kaczmirek L., (2008), "Prenotification in Web Based Access Panel Surveys. The Influence of Mobile Text Messaging Versus E-Mail on Response Rates and Sample Composition", *Social Science Computer Review*, 26 (2), 213-223.

Bryman A. (2004), *Social Research Methods*, Second Edition, Oxford University Press, Oxford.

de Bruijne M., Wijnant, A. (2013), "Comparing survey results obtained via mobile devices and computers: An experiment with a mobile web survey on a heterogeneous group of mobile devices versus a computer-assisted web survey", *Social Science Computer Review*, 31, 482-504, doi:10.1177/0894439313483976.

Callegaro M., Lozar Manfreda K., Vehovar V. (2015), *Web Survey Methodology*, London, Sage Publications Ltd.

Dillman D.A. (1978), *Mail and telephone surveys. The total design method*, New York, John Wiley.

Dillman D.A., Smyth J.D., Christian Melani M. (2009), *Internet, Mail, and Mixed-Mode Surveys: The Tailored Design Method*, 3rd ed. New York, John Wiley.

Fricker S., Tourangeau R., (2010), "Examining the Relationship between Nonresponse Propensity and Data Quality in Two National Household Surveys", *Public Opinion Quarterly*, 74 (5), 934-955.

Fuchs M. (2012), "Nonresponse and Panel Attrition in a Mobile Phone Panel Survey", Paper presented at the Federal Committee on Statistical Methodology Research Conference 2012, Washington.

Groves R.M., Lyberg L. (2010), "Total Survey Error. Past, present and future", *Public Opinion Quartely*, 74 (5), 849–879.

Guidry K.R. (2012), "Response quality and demographic characteristics of respondents using a mobile device on a web-based survey", Paper presented at the Annual Meeting of the American Association for Public Opinion Research, Orlando, United States (17-20 May 2012).

Keusch F. (2015), "Why do people participate in Web surveys? Applying survey participation theory to Internet survey data collection", *Management Review Quarterly*, 65, 183-216.

Lugtig P., Toepoel V., 2015, "The Use of PCs, Smartphones and Tablets in a Probability Based Panel Survey. Effects on Survey Measurement Error", *Social Science Computer Review*.

Lozar Manfreda K., Bosnjak M., Berzelak J., Haas I., Vehovar V. (2008), "Web Survey versus Other Survey Modes – A Meta-Analysis Comparing Response Rates", *International Journal of Market Research*, 50 (1), 79-104.

Lynn P., Kaminska O. (2013), "The impact of mobile phones on survey measurement error", *Public Opinion Quarterly*, 77, 586-605, doi:10.1093/poq/nfs046.

Mavletova A. (2013), "Data quality in PC and mobile web surveys", *Social Science Computer Review*, 31, 725-743, doi:10.1177/0894439313485201.

Mavletova A., Couper M.P. (2013), "Sensitive topics in PC web and mobile web surveys: Is there a difference?", *Survey Research Methods*, 7, 191-205.

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.

Maxl E., Haring W., Tarkus A., Altenstrasser M., Dolinar M. (2010), "Effects of Mobile Web Survey Invitation Modes on Non-Response", *International Journal of Mobile Marketing*, 5 (1), 5-14.

Millar M.M., Dillman D.A. (2011), "Improving Response to Web and Mixed-Mode Surveys", *Public Opinion Quarterly*, 75 (2), 249-269.

Sala E., Respi C., Decataldo A. (2014), "The use of text messages in a web-survey. The case of a survey of Italian graduates", Paper presented at the 7th Internet Survey Methodology Conference, Bolzano, Italy (1-3 December, 2014).

Sala E., Respi C., Decataldo A. (2015), "The use of text messages in a web-survey. The case of a survey of Italian graduates", Paper presented at the 6th Conference of European Survey Research Association – ESRA, Reykjavik, Iceland (13-17 July, 2015).

Shadish W.R., Cook T.D., Campbell D.T. (2002), Experimental and quasi-experimental designs for generalized causal inference, Mifflin and Company, Houghton.

Steeh C., Buskirk T.D., Callegaro M. (2007), "Using Text Messages in U.S. Mobile Phone Surveys", *Field Methods*, 19 (1), 59-75.

Toepoel V., Das M., Van Soest A. (2009), "Design of web questionnaires: The effects of the number of items per screen", *Field Methods*, 21, 200-213, doi: 10.1177/1525822X08330261.

Tolonen H., Aistrich A., Borodulin K. (2014), "Increasing health examination survey participation rates by SMS reminders and flexible examination times", *Scandinavian Journal of Public Health*, 42 (7), 712-717.

Tourangeau R., Rips L.J., Rasinski K.A. (2000), *The psychology of survey response*, Cambridge, New York, Cambridge University Press.

Virtaken V., Sirkiä T., Jokiranta V. (2007), "Reducing nonresponse by SMS Reminders in Mail Surveys", *Social Science Computer Review*, 25 (3), 384-395.

Wells T., Bailey J.T. Link, M. (2014), "Comparison of smartphone and online computer survey administration", *Social Science Computer Review*, 32, 238-255, doi: 10.1177/0894439313505829.