EVALUATION OF CHANGING IN RESPONDENTS' PARTICIPATION IN THE SURVEYS OF INFORMATION AND COMMUNICATION TECHNOLOGIES USAGE IN ENTERPRISES (ICT) AND RESEARCH AND DEVELOPMENT FOR BUSINESS ENTERPRISES (R&D)

Claudio Ceccarelli, Gabriella Fazzi, Samanta Pietropaoli[1]

1. Introduction

This paper describes a longitudinal analysis of response rate, response burden, and paradata for the survey of Information and Communication Technologies usage in Enterprises (ICT survey, 2014 - 2020 editions) and the Research and Development for Business Enterprises (R&D survey, 2017 - 2021 editions).

The purpose of this paper is to explain how the characteristics of surveys (i.e., number of questions, number of hard or soft prompts) and the characteristics of respondents (i.e., number of employees, enterprises involved in the sample of each survey edition) can affect the response rate. In section 2, the characteristics of the survey and the web application developed for the questionnaire are focused. Using paradata, the characteristics of the survey and the connection with response burden are described in section 3, a set of indicators describing the behavior of the respondent and non-respondent companies is presented, and a logit model is proposed, trying to explain the propension to answer. Finally, Section 4, some conclusions are drawn, suggesting how to reduce the burden on respondents and increase participation.

2. Data

The Italian Survey on information and communication technology (ICT) usage is carried out by Istat on annual basis on active enterprises in industrial and services sectors, with at least 10 employees[2]. The Survey on Research and Development for

[1] This article is the result of the collaboration between the authors. In particular: Introduction and paragraph 1 are attributed to Claudio Ceccarelli, paragraph 2 to Gabriella Fazzi and paragraph 3 to Samanta Pietropaoli.

[2] The principal aim of this survey is to supply users with indicators on information society: Internet activities (web site, social media, cloud computing) and connection used (fixed and mobile broadband), e-Business (use of software as ERP, CRM), e-Commerce, ICT skills, e-Invoice.

Business Enterprises (R&D) is also carried out by Istat and is an annual census of the population of the Italian companies that can be identified as "potential R&D performers" in the reference year[3]. ICT and R&D are both web surveys; the questionnaires are carried out on the Istat Business Portal[4] and designed with GX (Generalised Italian Data Collection System XML), a software tool for developing/designing electronic questionnaire, since 2014 and 2017, respectively.

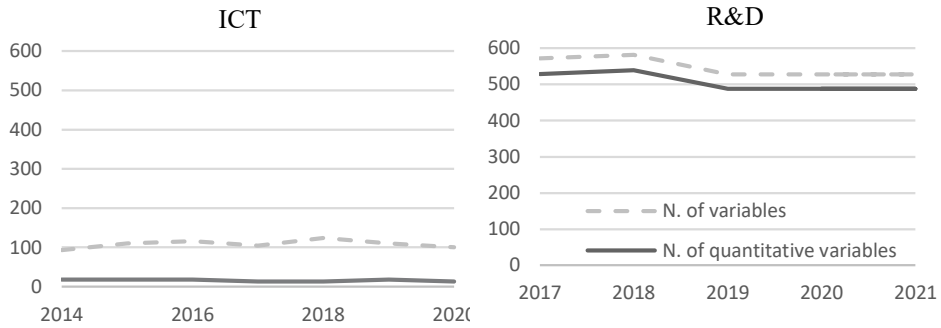
Paradata and metadata collected by the GX tool and the Business Portal starting from 2014 are analysed and we used them to try to explain the main factors for the response /nonresponse behaviour. Data on the number of requested variables, number of quantitative requested variables, number of hard and soft implemented prompts are used. The enterprises involved in the sample by size (number of employees) can classified and we are able to track whether the enterprise has done some action on the web Portal (answering the analysed surveys or other actions).

As shown in Figure 1, the ICT collects mostly qualitative information, contrary to the R&D, which collects mainly quantitative data. Also, the number of collected variables is different, in the ICT there are about one hundred for each edition, while in the R&D we work with 600 variables.

[3] It collects data about intramural and extramural R&D expenditure either within the unit (intramural) or outside it (extramural) and R&D personnel broken down by the business enterprise. Other information collected relates to the sources of funding for R&D activities and the type of research carried out (basic or applied research, experimental development).

[4] The Istat Business Portal is an integrated system for the management of data collection processes, which portal is at the same time an attempt to streamline the organization and production processes of business surveys.

Figure 1 – *Number of all type of variables vs number of quantitative variables in ICT and R&D surveys.*



Source: our elaborations on Istat data.

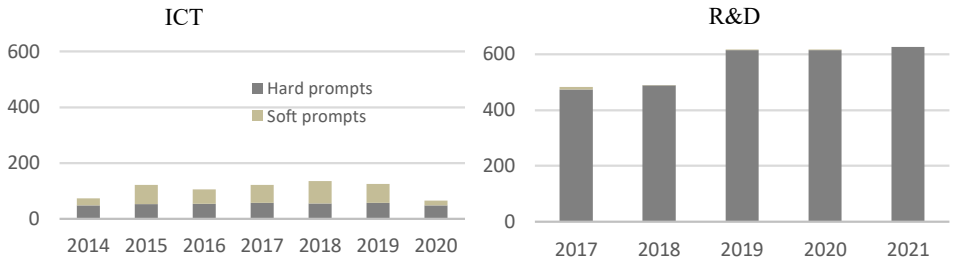
This characteristic, showed in Figure 1, highlights the different complexities of surveys: R&D is more demanding than ICT because respondents must find accounting information, referring to the current year and the two previous years, as well as expenditure forecasts for the following year.

The presence of many quantitative variables has led, over the years, to the insertion of many prompts, to allow for a consistent and higher quality data collection. In both surveys prompts are provided for unanswered questions, to check the question format and allowed character, and above all for consistency validations[5].

Nearly all prompts in the R&D survey are hard prompts, and the respondent cannot submit the questionnaire without correction. As Callegaro *et al*. (2015) note, "since validation messages interfere with the respondent's completion of the surveying task, they may be considered intrusive and annoying".

[5] The control of the consistency of responses with other data from the same survey or with the same answer in a previous survey of the same respondent.

Figure 2 – *Number of consistency validation prompts – ICT – R&D survey editions 2014-2020.*

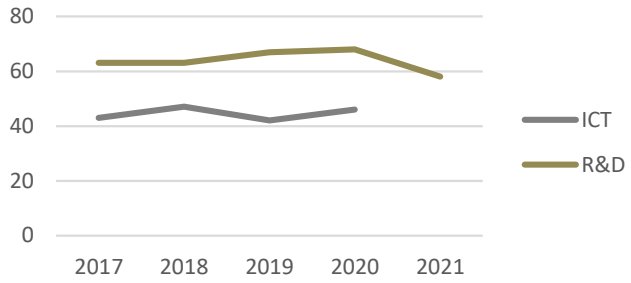


Source: our elaborations on Istat data.

Differences in the number of pieces of required information and constraints on completion lead to very different completion times for the two questionnaires as shown in Figure 3. On average it takes between 40 and 50 minutes for ICT online compilation, and usually more than 60 minutes for R&D. The commitment is heavier for larger companies, which fill in all sections of the questionnaire.

It is also worth mentioning some peculiar aspects of the response process in economic surveys, where respondents answer questions as representatives of their businesses and should be familiar with tables, matrices, and numerical information. Moreover, these surveys may need to be completed by multiple respondents, and the release of data may require approval by the company. Finally, printouts of web questionnaires are frequently used to support the preliminary process of identifying what information needs to be provided. Respondents often use paper forms as rough drafts before attempting to enter the data and answer the sequence of questions on multiple topics that appear on successive screens of a web survey (Morrison *et al*., 2008).

Figure 3 – *Compilation time online (mean in minutes) for ICT and R&D surveys.*



Source: our elaborations on Istat data.

In the ICT survey less than half of the questionnaire has been filled-in by only one referent. In one out of four cases more than two respondents had been involved in the retrieving of information (Table 1).

Table 1 – *People involved in ICT and R&D response process.*

	ICT 2020	R&D 2021
Single respondent	41.7 %	61.4 %
Two respondents	34.0 %	28.6 %
More than 2 respondents	24.3 %	10.0 %

Source: our elaborations on Istat data.

Table 2 – *ICT and R&D online and offline compilation time.*

	ICT		R&D	
	online	offline	online	offline
< 15 m	14.8	12.8	20.4	35.4
15-30 m	30.0	30.9	16.2	22.9
30-60 m	34.9	30.0	26.4	29.2
1-2 h	17.7	15.4	26.2	10.3
2-3 h	2.5	6.3	10.7	1.6
+ 3 h	0.0	4.5	3.1	0.6

Source: our elaborations on Istat data.

The time, and business cost, to consider for understanding the impact that completing the ISTAT questionnaires has on the business cannot just be the one automatically recorded by the systems. We need to add the time spent offline[6] to retrieve information or to contact other colleagues.
Comparing offline and online activities for the two surveys (Table 2), we observe that one out of four enterprises spend more than 1 hour offline working on the questionnaire for the ICT survey, compare with 12.5% for the R&D survey.

3. Response Rate Analysis

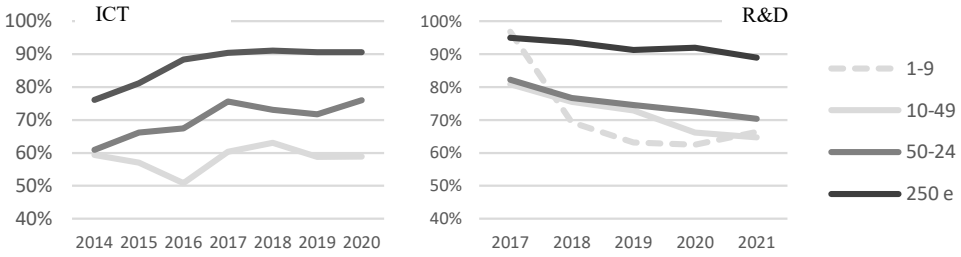
Focusing on the missing answers, we will try to understand why one enterprise out of three decides not to participate in the survey. The decision may depend on several factors: availability of the required information, lack of interest in the topics covered in the survey, authority attributed to the proposing agency, perceived response burden. Some firms (especially larger ones, with more than 250 employees) may be asked to respond to numerous surveys throughout the year, and this statistical

[6] Online time compilation, linked to paradata, is exhaustive and precise, which means it is automatically available for all enterprises who fill in the questionnaire, while data coming from the final section of the questionnaires are partial and subjective, as answering to this section was not compulsory.

burden is often significant. Understanding the incidence, but also the reasons for non-response can be important for identifying targeted strategies that can reactivate business involvement and increase participation.

The response rate trends among companies of different size ranges shows some interesting aspects: first, the response rate is always higher in larger companies; this is true not only for companies with more than 250 employees, which are subject to sanctions in the event of non-response, but also in companies that are not subject to sanctions.

Figure 4 – *Response rate for enterprise dimension (numbers of employees)- ICT and R&D.*



Source: our elaborations on Istat data.

This different response behavior may be due: a different strategy of reminders (the larger companies are also invited by telephone or by mail to fill out the questionnaire through a company that manages the outbound of respondents) or a more efficient internal organization in the larger companies, which allows time and/or resources to be dedicated to the filling out of the ISTAT questionnaires. It is interesting to observe the increases that in the ICT survey the response rate has had from 2017 onwards, and the decrease in the R&D survey (Figure 4).

We have examined some indicators that may provide some information regarding the phenomenon of missing responses:

- % of respondents who do not respond but had responded to the previous wave of the questionnaire;
- % of respondents who do not respond to a survey but respond to other surveys;
- % of non-respondents who have never accessed the Istat Business Portal.

In both ICT and R&D surveys, a portion of the sample must answer the questionnaire every year, while another portion changes from year to year. Those who had already responded to the survey the previous year are familiar with the questionnaire have already activated the information retrieval process within the company, and there are therefore somewhat facilitated to respond. On the other hand, returning to the sample each year increases the perception of statistical burden.

Using longitudinal data on sample inclusion and survey participation, we calculated for the two surveys the percentage of sample overlap (how much of the sample in *year t* was also present in *year t-1*) and the percentage of "lost" firms, i.e., missing firms that had responded the previous year.

Table 3 – *Samples overlap and "lost" enterprises in ICT and R&D survey.*

	ICT		R&D	
	Samples overlap	Lost	Samples overlap	Lost
2015 on 2014	32.0 %	10.9%	-	-
2016 on 2015	46.0 %	12.8%	-	-
2017 on 2016	33.0 %	6.1%	-	-
2018 on 2017	48.0 %	7.4%	68.0%	11.2%
2019 on 2018	62.0 %	9.3%	57.0%	11.7%
2020 on 2019	34.0%	9.4%	70.0%	6.8%
2021 on 2020	-	-	77.0%	5.3%

Source: our elaborations on Istat data.

In R&D surveys the sample overlap has been always bigger than that in the ICT survey, and the percentage of "lost" respondents is always lower. It is as if the R&D survey can count on a higher rate of loyalty from its respondents.

4. Focus on missing answers

As mentioned earlier, in both surveys approximately 66% responded to the questionnaire and completed the submission. What prevented the remaining 33% from fulfilling this task? A small percentage, 1.2% in ICT and 4.3% in R&D attempted completion but then left the questionnaire in draft form (Table 4).

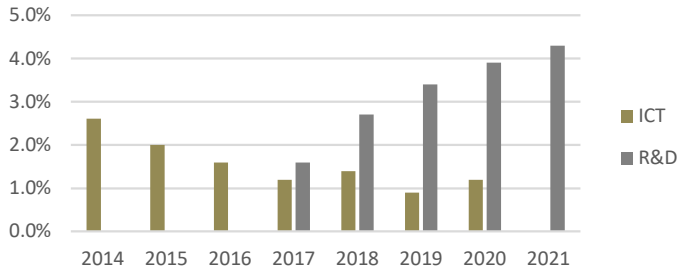
Table 4 – *ICT and R&D respondents behavior in 2021 and 2020 respectively.*

	ICT	R&D
Answered	66.1 %	66.8 %
Questionnaire in draft	1.2 %	4.3 %
Questionnaire opened and blank	3.4 %	4.1 %
Questionnaire never opened	11.9 %	15.8 %
Business Portal never opened	17.3 %	9.0 %

Source: our elaborations on Istat data.

Comparing the two surveys, we observe that the number of questionnaires left in draft has decreased over the years for ICT, while it has progressively increased for R&D (Figure 5). Every year, approximately 50% of the draft questionnaires (839 in 2021), moreover, had at least one active blocking error, that prevented their definitive sending. These are presumably questionnaires on which the high number of hard prompts had a discouraging effect.

Figure 5 – *Percentage of draft questionnaires in ICT and R&D surveys.*



Source: our elaborations on Istat data.

In fact, each company attempted the questionnaire twice before succeeding for ICT and 3.2 times for R&D. The multiplicity of attempts is due to the reporting of blocking errors for consistency checks that are triggered only at the end of the compilation, and that requires going back and correcting inconsistent data. In this regard, it is worth noting a hat is a red flag for R&D: 1172 companies (4% of respondents) changed their answer to the first question on R&D activity from "I did, I do, or I am planning to do R&D" to "I did not, I do not, and I am not planning to do R&D". This first response acts as a filter for the entire questionnaire, and a negative response allows to skip all the questions and the prompts, for a direct submission without penalty. To increase the participation of these companies, and enable them to complete the compilation, it is important to improve questionnaire design and usability principles (Couper, 2008).

Table 4 shows another interesting data point: 3.4% of the ICT sample and 4.1% for R&D merely opened the questionnaire, without beginning the compilation. To encourage their participation, therefore, it could be necessary to focus on more precise and clearer reminders. A considerable part of the sample, however, has never even opened the questionnaire webpage: this is 29.2% for ICT and 24.8% for R&D. Of these, a part ignored the invitation to access the questionnaire, yet accessed the Business Portal. They may have had difficulty in locating the web page where to begin the compilation, or they have other questionnaires to fill in and forgot or decide not to answer this one. An analysis of the user experience of the Portal could help to better understand the reasons for this dispersion, which led to the loss of 12% of the ICT sample and 16% of the R&D sample in the last edition of the survey.

The last interesting and numerically relevant category of non-respondents are those that we will call "wanted", that is, companies that have never even accessed the ISTAT Portal. These enterprises (17.3% for ICT and 9% for R&D) simply ignore the invitations and reminders they have received or have had difficulty accessing the Portal, which has discouraged them from participating in any economic survey.

5. Propension to answer: a logistic model

The data analysis concerns almost thirty-nine thousand enterprises included in the R&D sample for the year 2021 and twenty-one thousand enterprises from the ICT sample for 2020.

For each sample we studied the propension to answer through a logit model, with the dependent variable being the response registered and explanatory variables are some characteristics of the enterprises. We studied the influence of monetary sanction affecting enterprises with large dimension (over 250 employees)[7], the presence of the enterprise in previous wave of the questionnaire, and in all editions - as an indicator of low and medium statistical burden in the year (see Table 5). The indicator of statistical burden has been calculated with reference to the number of Istat statistical surveys in which the enterprise is required to participate in the year: low - if the enterprise is involved in less than five surveys; medium - if it is involved on at least five and less than twelve surveys; high – if it is involved on at least twelve surveys. In the model, our reference variable is the high burden.

Table 5 – *Variables used in the logit model.*

Variable	Description
RESPONSE	1-Respondent; 0- otherwise
SANCTION	1- Enterprise with 250 employees and over, 0- otherwise
PREC_EDITION	1- Enterprise included also in the sample of the previous years 0- otherwise
ALL_EDITION	1- Enterprise included also in the sample of all the previous years 0- otherwise
LOW_BURDEN	1- low statistical burden, 0- high statistical burden
MEDIUM_BURDEN	1- medium statistical burden, 0- high statistical burden

6. Results

Table 6 reports the change in model fit to the data obtained by omitting one explanatory variable at a time and instead retaining the others (Chi-square statistic), with the p-value associated. The results for both the survey show that each of the variables has a significant effect on the propensity to answer of the enterprise (column 3 and 5); in particular, the level of the Chi-Square statistic for both survey (column 2 and 4) shows that having participated in the previous edition is the most relevant factor among those considered. This is followed by monetary sanction,

[7] All enterprises in the sample are informed about the mandatory nature of the response, however, the sanction only affects large companies.

while, low level of burden, participation in all precedent editions seem to have a more limited impact on the phenomenon under consideration.

Table 6 – *Explanatory variables for Wald Chi-square statistic and associated p-value.*

	ICT		R&D	
	W.Chi-Sq.Test	Pr > Chi-Sq.	W. Chi-Sq.Test	Pr > Chi-Sq.
Sanction	69.7865	<.0001	65.6011	<.0001
Pre_edition	95.3255	<.0001	79.5082	<.0001
All_edition	36.7653	<.0001	0.6554	<.0001
Low_burden	41.2861	<.0001	4.6974	<.0001
Medium_burden	24.1219	<.0001	2.9872	<.0001

Source: our elaborations on Istat data.

When all variables involved in the model are categorical, as it is the case here, interpretation of the results may be more straightforward by directly examining the Odds Ratio[8], rather than the parameters.

Table 7 – *Odds Ratio for model-related regressors.*

	ICT	R&D
	Estimate	
Sanction (1 vs 0)	1.826	2.089
Pre_edition (1 vs 0)	1.361	39.445
All_edition (1 vs 0)	1.454	1.411
Low_burden (low vs high)	0.039	2.982
Medium_burden (medium vs high)	0.084	2.386

Source: our elaborations on Istat data.

All coefficients of the equation have significant values at 1% significance level. These coefficients indicate the change of the log odds of being respondent at a unit increase of the predictor variable.

Among the principal results in the ICT survey, the monetary sanction appears to be the main factor that drives companies to participate, as it corresponds to the highest probability of being a respondent. The burden of the company in the same year and the involvement in the previous edition of the survey seem to have less weight. In R&D, however, it is precisely this element that strongly encourages participation: this can be explained by the fact that the survey also requires retrospective data each year, and so for an enterprise that has already responded to edition *t-1*, the task of responding to edition *t* is, in effect, much simplified.

[8] An odds ratio (OR) is a measure of the association between an exposure and an outcome. The OR represents the probability that an outcome will occur given a particular exposure, compared with the probability that the outcome will occur in the absence of that exposure.

7. Discussion and future developments

The results of the analysis of the paradata and metadata of the two surveys are obviously not exhaustive and open the field to new ideas for analysis and research. A specific focus on the "wanted", that is on the "hard-core" of companies that have never accessed the Portal, is necessary to study strategies aimed at obtaining their collaboration, capturing their attention. Again, multivariate analysis can be conducted on a broader and longitudinal database, to try to understand the factors that lead the two surveys to have opposite trends in the response rate.

However, the main findings presented in the paper are already useful to indicate some actions that the Data Collection Directorate can take to try to reduce the burden on the companies themselves, without losing the wealth of information collected and hopefully increasing response rates and data quality.

It would be possible to try to engage the wanted with targeted outbound call campaigns, guiding them through their first access to the Portal. The internal paths within the Portal itself can be redesigned, improving the user journey, and making the call to action more immediate: highlighting the deadline of each survey and making access to the questionnaire immediate can be two first steps in this direction, improving the user experience and usability of the online data collection instruments and web Portal.

The questionnaires need to be as clear as possible, and the data consistency check required some simplification.

The burden would be reduced if different questionnaires, from different surveys, were linked together and with the administrative data that the enterprise already communicates to the PA. A firm involved in several samples would not be forced to provide the same information multiple times. This requires a work of standardization of variables and coordination of survey times between different surveys that should be discussed at Eurostat tables (EUROSTAT, 2014).

Finally, it would be important that the primary motivation for responding to the proposed questionnaires was not fear of monetary sanction. The strategies to be put in place to engage these respondents cannot be just "punitive", for example, by focusing on economic sanctions. It would be opportune to focus on more effective communication strategies and a return of information on the respondent capable of making them perceive the importance of the data collected by the Institute even for the individual business. An example is the experience of Portugal, which returns sectoral benchmark reports to companies that participate in the surveys.

For companies to understand the importance of the data collected by official statistics for the country, it is necessary that those same data return part of their value to the companies themselves.

References

CALLEGARO M., LOZAR MANFREDA K., VEHOVAR V. 2015. *Web survey methodology*. London: Sage.

COUPER M. P. 2008. *Designing effective web surveys*. New York: Cambridge University Press.

EUROSTAT 2014. *Memobust handbook - Handbook on Methodology of Modern Business Statistics.*

MORRISON R. L., STOKES S. L., BURTON J., CARUSO A., EDWARDS K. K., HARLEY D., HOUGH C., HOUGH R., LAZIRKO B. A., PROUDFOOT., S. 2008. *Economic Directorate guidelines on questionnaire design.* U.S. Census Bureau.

SUMMARY

Evaluation of Changing in Respondents' Participation in the Surveys of Information and Communication Technologies Usage in Enterprises (ICT) and Research and Development for Business Enterprises (R&D)

In the paper we analyzed paradata and metadata of two business surveys: ICT or R&D survey, with the aim to understand the response behavior of the respondents. Starting with a descriptive analysis, we focus on characteristics of the surveys (i.e., number of questions, number of hard or soft prompts) and on the characteristics of the respondents (i.e., number of employees, enterprises involved in the sample in each survey edition) for both the surveys. Then we analyzed response rate by enterprise size, and we propose a set of indicators describing the behavior of the respondent and non-respondent enterprises. We also study paradata of the users as the time taken to compile the questionnaire, online and offline, with the goal to find some actions to improve the participation of the respondents and their perception of the official statistics. For this reason, we also focused our attention on a specific subset of respondents - the so-called "wanted" - the ones who have never answered the ICT or R&D survey or any other Istat survey- and the “lost” ones, which had answered to the previous year but are missing in the current year.

Finally, we apply a logit model intending to explain which of respondents’ characteristics affect response rate and define what kind of actions can be taken to improve participation.

Claudio CECCARELLI, Istat, clceccar@istat.it
Gabriella FAZZI, Istat, fazzi@istat.it
Samanta PIETROPAOLI, Istat, pietropaoli@istat.it