

MONITORING PROCESS AND NON-SAMPLING ERRORS CONTROL IN THE IT-LFS SAMPLE SURVEY¹

Serena Capuano, Francesca Carta, Alessandro Martini, Alessia Sabbatini

Abstract. This study presents strategies to minimise non-sampling errors in the IT Labour Force Survey. Emphasis is placed on the rigorous monitoring of interviews and processing data to ensure the reliability and accuracy of the collected information. Ongoing data analysis is conducted to promptly identify and address inconsistencies or biases, thereby enhancing data quality. In the context of the LFS, this takes on a very high value due to the complexity of the defining framework and the breadth of the concepts involved.

Several strategies are developed to ensure these objectives: i) instructional notes and ad hoc training sessions are systematically implemented to enhance the proficiency of interviewers and ensure standardised data collection procedures; ii) real-time remote monitoring of CATI interviews: listening to the conversation between the interviewer and respondent, as well as viewing the interviewer's screen in real-time, through remote tools; iii) on-site monitoring: ISTAT staff periodically visit the interview sites to observe interviews, ensuring adherence to quality standards directly; iv) post-interview monitoring and data analysis to evaluate the logical flow of the questionnaire as well as the consistency of the collected information with the data time series.

To assess the impact of interviewers on the quality of survey results, a synthetic indicator was developed to highlight areas significantly influenced by interviewer effects (as measured by the intraclass correlation coefficient), support corrective measures through evaluations at the individual-interviewer level, and plan targeted training activities.

These strategies help identify and correct potential sources of non-sampling error in relevant process parameters and ensure a thorough evaluation of both methodological and content-related aspects of the interviews, enabling timely interventions when necessary.

1. Monitoring process and non-sampling error control in the IT-LFS sample survey

The work presents strategies adopted to identify and minimise non-sampling errors in the IT Labour Force Survey during and following the transition to the new

¹ For research articles with several authors, a short sentence specifying their individual contributions can be provided here. The affiliations and contacts of the authors should NOT be reported here, but at the bottom of the last page.

companies conducting field interviews using CAPI and CATI techniques. The outsourcing of CAPI and CATI interview services to new providers was conducted through a public procurement process, in accordance with the Public Contracts Code, which governs the awarding and management of public tenders in Italy. The change involved implementing an electronic questionnaire, establishing new procedures and training interviewers.

Non-sampling errors encompass all sources of error that are independent of the sampling process. They can occur at any stage of the survey process and are primarily associated with data collection and processing. They arise from multiple factors (inaccurate respondent-reported data and missing observations mainly due to misleading definitions and concepts, inadequate frames, unsatisfactory questionnaires, defective data collection methods, tabulation and coding, incomplete coverage of sample units, etc.). Non-sampling errors can be categorised into two distinct types: random errors and systematic errors. The former are stochastic variations whose aggregate effects tend to neutralise when the sample size is sufficiently large. Systematic errors are directionally consistent, accumulate across the dataset and introduce bias into the results. They represent the primary challenge in ensuring the quality of survey data. Non-sampling errors can be further divided into errors of observation and errors of non-observation. The first category encompasses measurement inaccuracies stemming from various sources, including the method of data collection, interviewer influences, deficiencies in the measurement instrument (i.e., questionnaire deficiencies), respondent-related factors, post-survey data processing, and alteration of reported data. These errors introduce potential bias into the collected data and require methodological controls to mitigate their impact, as they involve differences between reported/recorded values and the “true” or underlying value. Errors of non-observation include coverage and nonresponse errors.

Some conditions remained unchanged when responsibility for field data collection was transferred to new companies. Therefore, our analysis has focused on specific types of error (Table 1). For example, we know that errors could not be found in misleading definitions and concepts, unsatisfactory questionnaires, tabulation, and coding, since these aspects remained stable and were, in some cases, governed by automatic procedures that were under the control of ISTAT. Also, the questionnaire includes internal consistency checks, and the collected data are verified for consistency and coherence. The strategies adopted to minimise non-sampling errors in the IT-LFS only address some error types.

Concerning observation errors, we did not consider specification errors, which occur “when the concept implied by the survey question differs from the concept that should be measured in the survey”. This is because there has been no change to the conceptual and technical framework with the new companies. At the same time,

we did not account for processing errors that occur after data collection and during processing, as the way data are recorded, coded, stored, or analysed remains unchanged. We consider the situation stable compared to the past (e.g., software bugs in data processing, mistakes in data transformation or aggregation, etc.).

Table 1 - *Types of non-sampling errors identified and monitored, and actions taken.*

Types	Sub-types	Error occurrence	Monitoring carried out	Action
Coverage errors: omissions (undercoverage), erroneous inclusions and duplications (overcoverage), and misclassifications of units in the survey frame		NO	-	-
Measurement error: difference between measured values and true values caused by inefficiencies in the questionnaire, interviewer, respondent or survey process	Poor questionnaire design: misleading or confusing questions	NO/Unchanged	-	-
	Interviewer bias: inexperienced or poorly trained interviewer or displaying incorrect or inappropriate behaviour	YES	YES	Remote and on-site listening, data analysis
	Respondent error: providing incorrect answers, having faulty recollections, exaggerating or downplaying events, and giving socially acceptable answers	NO/Unchanged	-	-
	Survey process problems: proxy responses, lack of control over survey procedures	YES	PARTIALLY	Data analysis
Non-response error: difference between the estimates that would have been obtained in the absence of non-response and those that are observed (or inputted)	Total nonresponse error: data for a sampling unit are missing (respondent was unavailable, temporarily absent, unable to participate, or refused to participate in the survey)	YES	YES	Remote and on-site listening, data analysis
	Partial nonresponse error: respondents provide incomplete information (the questions are difficult to understand or because they refuse or forget to answer a question)	Not relevant, mandatory answers	-	-

Table 1 (cont.) - Types of non-sampling errors identified and monitored, and actions taken.

Types	Sub-types	Error occurrence	Monitoring carried out	Action
Processing error: errors in the capture, coding, editing and tabulation of data, and in the assignment of survey weights	Coding errors: the same answer is coded differently by different coders due to poor training, incomplete instructions, variance in coder performance, data entry errors, or machine malfunction	YES	YES	Remote and on-site listening
	Data capture errors: data are not entered as they appear in the questionnaire	No/Unchanged + automated data capture method	-	-
	Errors in editing and imputation: poor quality of the original data or by its complex structure	No/Unchanges	-	-

Source: Our elaboration based on Statistics Canada, <https://www150.statcan.gc.ca/n1/edu/power-pouvoir/ch6/nse-enda/5214806-eng.htm>.

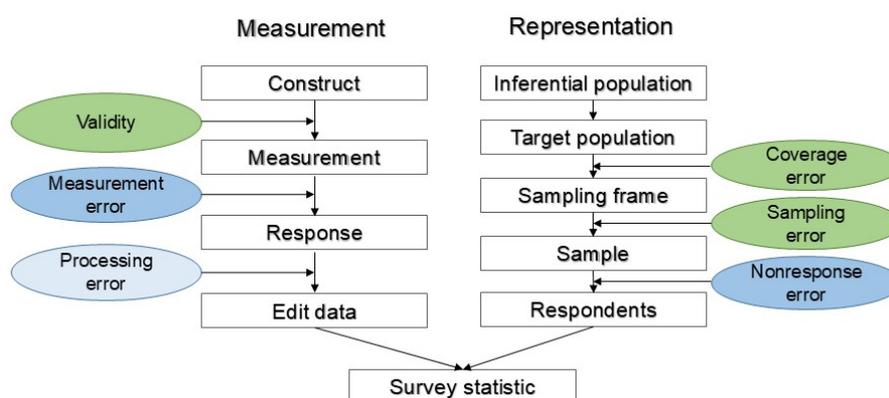
Instead, consistent efforts were made to analyse measurement error, the deviation between observed and true values. This occurs when data are inaccurately solicited, reported, received, or recorded, and is strictly influenced by the interviewer's behaviours. In this context, we also investigated the use of proxy interviews, given the interviewer's tendency to use proxies, which introduces additional uncertainty and classification errors.

We also checked for nonresponse errors, which occur when discrepancies arise between the survey statistics and the frame population. We focused on total nonresponse error (unit nonresponse) in particular, as it systematically affects all survey-based estimates. This is one of the most critical sources of error, as it can bias population estimates across subgroups.

The diagram in Figure 1 illustrates the key sources of error in statistical surveys, classified into *measurement errors* (e.g., response and validity issues, processing errors) and *representation errors* (e.g., coverage, sampling, and nonresponse errors). These errors can significantly affect the accuracy of survey estimates. To ensure reliable results, the aim is to minimise them through targeted methodologies and quality control measures. The strategies implemented involve a combination of methodological approaches and control tools. The primary focus is on measurement errors, followed by nonresponse errors, and to a lesser extent, processing errors

related to coding (Figure 1). These strategies have been developed and are continuously applied to prevent, identify, and address these challenges.

Figure 1 - Total survey error components linked to steps in the measurement and representational process.



Source: Groves et al., 2004.

On the controlling side, the following are carried out:

1. Real-time remote monitoring of CATI interviews: listening to the interviewer and respondent's conversation and viewing the interviewer's screen in real-time through remote tools. The activities are carried out in compliance with the relevant regulations on personal data protection (EU Regulation 2016/679, Legislative Decree 196/2003, and subsequent amendments).

2. On-site monitoring: ISTAT staff periodically visit the interview sites to observe interviews, ensuring adherence to quality standards directly.

3. Post-interview monitoring and data analysis involve evaluating the logical flow of the questionnaire and the consistency of the collected information with the available historical data.

4. To identify the interviewer effect on the quality of the survey results, a synthetic estimate has been developed to highlight issues affected by a significant interviewer effect (intraclass correlation coefficient), and to make corrections to the process, an evaluation at the level of the individual interviewer (the rate of interviewers with an estimate significantly different from 0).

To counter and resolve anomalies or misconduct, instructional notes, ad hoc training sessions, and debriefings are systematically implemented to enhance interviewers' proficiency and ensure standardised data collection procedures.

2. Quality monitoring strategies in interview conducting

To minimise non-sampling errors in the conduct of interviews for the IT-LFS, continuous quality monitoring is carried out regarding the administration of interviews and the management of contacts with sampled households to identify potential weaknesses and needs for targeted training feedback. The following operational strategies are implemented:

- a. Real-time remote monitoring of CATI interviews.
- b. On-site monitoring of CATI interviews at the interviewers' workplace.
- c. Interviews and contacts were monitored ex-post through quality indicator analysis related to interview conduction.

The monitoring activities allow verification of interview administration methods and contact management (e.g., question comprehension, interaction difficulties with respondents, proper questionnaire administration, appointment scheduling, household substitutions), with the goal of timely intervention in case of problems, including targeted training initiatives and, at the same time, to monitor and verify the occurrence of proxy interviews.

a. *Real-time remote monitoring of CATI*: Interviews is conducted to assess the quality of telephone interviews using tools that enable real-time listening and viewing of CATI-mode interviews via a VPN connection to the interviewers' operational headquarters. The activities include:

- live listening of the conversation between the interviewer and respondent, with prior notification to the interviewer;
- real-time screen viewing of the application used by the interviewer.

The activities aim to follow interviewers to detect erroneous behaviours (e.g., suggesting or anticipating answers, misleading response categories). A weekly monitoring plan is structured in a matrix organised by days and time slots to ensure a fair distribution of listening activities among ISTAT colleagues, daily monitoring of interviews, observing the same interviewer at different times, and accounting for variations in interview style throughout the day and the entire week.

The listening session includes descriptive evaluation, addressing aspects such as question reading, coding correctness, refusal management, empathy, etc., and synthetic evaluation, rated on a scale from "insufficient" to "excellent". Evaluations are stored in an interviewer assessment registry to track individual performance over time. The platform enables interviewer selection and focused monitoring of specific questionnaire sections, allowing targeted checks based on emerging needs, such as active/inactive interviewers, interview duration with significant deviations from the average/standard, below-average qualitative performance, and anomalies in quality indicators. These listening sessions help assess question comprehension, identify

difficulties with wording or question structure, and address them through appropriate training feedback.

The remote monitoring platform is technically reliable, easy to use independently, secure in data and communication protection with the authentication into the switchboard dashboard, and compliant with privacy regulations. It has been implemented on laptops used for remote monitoring.

b. *On-site monitoring of CATI Interviews:* ISTAT periodically conducts control activities at interviewers' workplaces, including direct listening via dual headsets and remote verification of ongoing interviews (audio and video) via dedicated CATI stations. These controls comply with privacy regulations and provide detailed insight into interview quality and interviewer behaviour.

c. *Ex-post monitoring of contact and Interview conduct assessment:* Some paradata are also collected and monitored to improve data quality. Quality indicators and evaluations are used to assess the entire data collection process. Provisional and final contact outcomes are analysed, and specific indicators are generated:

- response rate, base family response rate, refusal rate, unavailability rate, average number of contacts required to complete household interviews;
- incorrect telephone contact rate;
- newly acquired telephone contact rate;
- interview distribution by survey day, time slot, and interviewer;
- average interview duration.

In this context, the checks also concern total nonresponse error in the case: 1) not-at-homes: prospective respondents who may not be at home when enumerators visit their households, controlling first/last interviews and request for differentiating time scheduling contacts, 2) refusals, respondents who refuse to give information for whatever reasons, monitoring data and possible increase.

3. Analysis of collected individual interview data

Frequency distributions, contingency tables, and indicators are developed based on key LFS variables. The post-interview monitoring and data analysis focused on:

- The respondent's response style, particularly the tendency to provide the same answer to a set of questions on a specific topic (response set on work satisfaction and working hours).
- The percentage of questions confirming the information provided in previous interviews, especially for specific questions or categories of respondents (for

example, 16-24-year-olds who confirm their educational qualification from the earlier interview).

- The consistency of specific key aggregates: percentage of persons employed, percentage of persons neither seeking nor available for work, and unpaid family workers. And, in general, significant percentage changes in key variables (reason for not seeking a job, differences in hours worked during the week compared to contractual working hours) and in questionnaire flow, shortening the pathway questions.

All results are observed and compared with the time series, assuming that in the past we obtained a “true value” (or better) of the characteristic under study for every unit in the population.

Survey data analysis proceeds in conjunction with remote listening, and the joint results are used during training to remedy interviewer shortcomings.

4. Evaluation of the interviewer’s effect

A hierarchical structure is evident in the LFS data: respondents are nested within interviewers, and in both CAPI and CATI modes, we find clusters of households (and individuals) interviewed by a particular interviewer. An appropriate class of models for analysing this type of data is the multilevel modelling approach, which accounts for variance due to differences between interviewers. In this paper, we cannot investigate in depth how interviewer characteristics may affect respondents’ answers and interviewer-respondent interactions, as additional information on interviewers is unavailable due to privacy issues. On the other hand, we are interested in providing a consistent measure of the “total” interviewer effect, which should be consistently monitored during data collection, and in what can be done to reduce this source of bias.

We have started applying it during the IT LFS data collection phase to identify and estimate the interviewer’s effect on survey quality. Moreover, only in the CATI mode is the random allocation of interviews to interviewers. CAPI interviewers collect data in a specific area, which often has a particular influence on the data. With the new data collection framework, we have unique interviewer codes for both techniques. We conducted this analysis only on the CATI subsample to obtain a reliable estimate of the interviewer effect, which was not confounded by territorial heterogeneity.

The response variables considered in our analysis are mainly dichotomous (activation of critical items or questionnaire filters, new items in the questionnaire, “Don’t Know- No Answer” responses to specific items, Proxy interviewing). For

these kinds of variables, the model we have chosen is the logit-based model described by:

$$y_{ij}|u_j \sim B(1, \pi_{ij}) \quad (1)$$

$$g(\pi_{ij}) = \mu + u_j \text{ with } u_j \sim N(0, \sigma_u^2) \quad (2)$$

$$\pi_{ij} = \frac{\exp(\mu + u_j)}{1 + \exp(\mu + u_j)} \quad (3)$$

$$\rho = \frac{\sigma_u^2}{\left(\frac{\pi^2}{3} + \sigma_u^2\right)} \quad (4)$$

π_{ij} denotes the probability of having the outcome 1 for the response variable y_{ij} and the j -th interviewer, and u_j , which are assumed to be independent identically distributed (IID) random variables $N(0, \sigma_u^2)$, are supposed to represent the variability due to the interviewer's effect. The intra-class correlation coefficient (ICC) is estimated as the ratio of group variance to total variance. It measures the degree of homogeneity within units in the same cluster and, in our context, represents the interviewer effect. We use a random intercept model to obtain a reliable estimate of the interviewer effect, as the introduction of additional explanatory variables makes the variance composition in (4) less precise. To monitor the data collection process, a synthetic estimate can highlight the topics affected by a significant interviewer effect. Still, we also need some "interviewer-level" evaluation to correct the process. For this reason, once a suitable model has been fitted, it is possible to estimate an interviewer-level effect using empirical Bayes prediction. Prediction procedures provide (\hat{u}_j) estimates for each interviewer with their relative confidence intervals, which allow us to analyse the interviewers' performance ranking and identify those with a predicted estimate significantly different from zero.

The interviewer effect has been tested across many aspects of the data, using the last four quarters for which the new operator is contracted to manage the interviews (Table 2). For most of the aspects investigated, the effect is not significant, indicating that the interviews are of average quality. The first important result is that, considering the respondent's gender as a target variable, a significant interviewer effect does not emerge, consistent with the literature.

Table 2 - Intraclass correlation (interviewer's effect) for specific variables in the IT-LFS questionnaire. Year 2024-2025.

Target Variable	2024			2025
	Q2	Q3	Q4	Q1
Respondent's gender	0.00	0.00	0.00	0.00
Filter to the previous work exp section	0.00	0.01	0.03	0.02
Lifelong learning indicator	0.07	0.18	0.18	0.06
Response set - work satisfaction	0.18	0.33	0.46	0.25
Confirm on educational level	0.00	0.04	0.06	0.03
Response set - working hours	0.06	0.07	0.08	0.04
Item nonresponse on income	0.09	0.25	0.30	0.26
Not seeking, not available	0.05	0.18	0.30	0.37

Source: IT-LFS data.

We have found a relatively low but significant value for ρ in predicting item activation from prior working experience: responding “No” allows for a shorter interview duration (i.e., skipping an entire questionnaire section). Some results also show that specific, intensive training sessions aimed at reducing the interviewer's effect in the last quarter have been successful, particularly for the lifelong learning indicator and the response set work satisfaction. Higher intraclass correlation coefficients indicate the interviewer's difficulties in handling sensitive topics; this is evident in the item nonresponse for the income item. In this case, the interviewer effect is high and steady over the last quarter of the period, ranging from 0.25 to 0.30. This confirms the great importance of interviewers' ability to create a climate that allows discussion of sensitive topics.

5. Findings and considerations

In 2006, a Eurostat Task Force was set up to review the quality of the EU-LFS along the dimensions of the quality framework for statistical output of the European Statistical System (ESS), identify weaknesses, and recommend improvements (Eurostat, 2009). The TF recommendations about “Interviewers and fieldwork organisation” represent clear reference points for the proposed work, in particular the following ones: a) provide interviewers with appropriate training on how to conduct the interview, covering both communication skills and survey content. Training should be provided continuously (R20); b) organise periodical focus groups and debriefing for interviewers to review and tackle critical issues (R21); c) instruct the interviewers in general not to adapt questions to the respondent but just to read them out as they are written in the questionnaire, as personal interpretations may lead

to biased results (R22); d) continuously monitor and systematically assess the quality of the fieldwork (R23).

The lack of trained and experienced field interviewers and of adequate field supervision can influence survey measurement in various ways. As has been widely demonstrated, the mere presence of an interviewer induces respondents to give more socially desirable answers than do self-administered questionnaires (Tourangeau and Yan, 2007). Beyond that, the interviewer's tasks may include contacting the respondent and obtaining co-operation, explaining the interview context, formulating questions and selecting and coding answers. Consequently, interviewers can influence many sources of error in a survey. Therefore, as interviewers can be considered a relevant source of error, in the current article, we presented strategies to minimise this impact on data, with a special focus on rigorous monitoring of interviews and processing data to ensure the reliability and accuracy of the collected data.

In the analysis, our intent is not to provide evidence that respondents can evade or cannot adequately differentiate between questions, but to examine the extent to which interviewers mediate these processes and try to counteract the corruptive effects of the data.

The process we implemented is continuous, and monitoring persists. Periodically, the company is informed with a report and required to take action if the interviewer's misconduct is repeated.

Starting from the idea that eliminating interviewer error is impossible, the aim here is to achieve data quality by giving the interviewer a more conscious, collaborative and proactive role to reduce the gap between the actual and collected data. Debriefings are organised for this purpose at various times throughout the year, as part of continuous quality improvement efforts.

Methodological advancements reflect the core of scientific investigation, where theoretical frameworks guide the development of a discipline. By carefully breaking down complex phenomena, researchers can better identify and understand the causal mechanisms at work. This process highlights the importance of taxonomies as vital tools for organising research efforts and strengthening analytical thinking. (Groves and Lyberg, 2010).

These strategies ensure a thorough evaluation of the interview's methodological and content-related aspects, allowing timely interventions when necessary. These activities help ensure proper interview administration, improve interviewer effectiveness and optimise the overall quality of the survey.

However, the extent of these critical issues, among other reasons, also indicates the survey's complexity, organisation and management (Banda, 2003).

References

- BANDA J.P. 2003. Nonsampling errors in surveys. *Expert Group Meeting to Review the Draft Handbook on Designing of Household Sample Surveys*, United Nations Secretariat ESA/STAT/AC.93/7, Statistics Division.
- EUROSTAT. 2009. Task force on the quality of the Labour Force Survey, *Final report*, Luxembourg, <https://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-RA-09-020>.
- GROVES R.M., FLOYD F.J., COUPER M.P., SINGER E., TOURANGEAU R. 2004. *Survey Methodology*. New York: Wiley.
- GROVES R.M., LYBERG L. 2010. Total Survey Error: Past, Present, and Future, *Public Opinion Quarterly*, Vol. 74, No. 5, pp. 849–879, <https://doi.org/10.1093/poq/nfq065>.
- TOURANGEAU R., YAN T. 2007. Sensitive questions in surveys, *Psychological Bulletin*, Vol. 133, No. 5, pp. 859–883.

Serena CAPUANO, ISTAT serena.capuano@istat.it
Francesca CARTA, ISTAT francesca.carta@istat.it
Alessandro MARTINI, ISTAT alessandro.martini@istat.it
Alessia SABBATINI, ISTAT sabbatini@istat.it