The challenges of introducing CAWI in the Portuguese LFS

Célio Oliveira [celio.oliveira@ine.pt], Statistics Portugal
Pedro Silva   [pedro.silva@ine.pt], Statistics Portugal

1   Introduction

The planned changes for the Portuguese Labour Force Survey in 2021, such as the introduction of the wave approach, the extension of dependent interviewing and the modifications of the questionnaire itself, open an opportunity to rethink the way we collect the information. Here we focus on the expected challenges in this regard (for discussion of other dimensions, such as series breaks, see Torres 2018), mainly promoted by the introduction of CAWI in the current mix of data collection modes (CAPI and CATI).

The aim of introducing CAWI is to reduce the burden of response for those who want or need a more flexible channel of response (time and place). At the same time, we expect answers from respondents who would not participate otherwise, providing representativeness gains. On the other hand, society is more used to interact with government entities using the web, and so we anticipate more demand from respondents to respond through CAWI.

CAWI in a multi-mode design is not new at Statistics Portugal. It was first introduced in 2011 Census, then in the European Health Interview Survey (EHIS) of 2014, followed by other surveys such as the survey on Information and Communication Technologies (ICT) usage in households and by individuals since 2018. This provides us with some experience on dealing with web interviews within a mixed mode approach, as well as the support of an existing web platform for data collection that is part of an integrated multi-mode IT support and management system.

However, in addition to being one of the most important operations (measured by media coverage and political discourse), the LFS has unique characteristics compared with previous CAWI experiences at Statistics Portugal that impose news challenges on how to implement CAWI and coordinate it with other modes. In the next sections, after an initial general overview of the Portuguese LFS, we outline some of the key dimensions for data collection and options being evaluated.

2   An overview of the Portuguese LFS (current and future)

The Portuguese LFS has a two-stage cluster sampling design of about 22,500 dwellings with a rotation scheme of 6 waves (quarters). The reference week is fixed with the collection period occurring in the following 2 weeks (most of them in the 1st week for the monthly estimates). Response is mandatory, and while the first interviews (mainly in the first rotation) are approached by CAPI, CATI is currently used as the preferred mode for subsequent interviews. The current response rate is approximately 70% with about 30% from CAPI and 70% from CATI.

In the future, the rotation scheme will be maintained with the first rotation exclusively collected through CAPI (corresponding to a longer interview, namely as a result of the wave approach), and with CATI and CAWI being offered as the main possibilities in the following rotations.

Regarding the questionnaire, in line with what already happens for CAPI/CATI, with the introduction of CAWI, we maintain a unique questionnaire being used for all three modes (a unimode design),

with slight adjustments in language when necessary. Its review and reformulation was, nonetheless, done bearing in mind the introduction of CAWI, for instance in attempts to simplify the questions and additional "help" information (common to the different modes).

3   CAWI and its challenges to the LFS multi-mode data collection

As mentioned, in contrast to our previous surveys involving CAWI, the LFS is a continuous operation with a fixed reference week and with a very short data collection period for each household (only up to 2 weeks). This leaves little space for corrections and imposes the need for greater thoroughness in its preparation and management.

In addition, the longitudinal dimension is of crucial importance, namely regarding the need for the same individual to keep its ID through time, which raises some confidentiality concerns for showing previously provided information in CAWI (given a sample of private dwellings).

In this section we present some of the main challenges associated with CAWI in this context, and the options under consideration.

3.1   Multi-mode organization including CAWI

3.1.1   CAWI for whom?

As previously stated, web interviews will only be an option after the first rotation. In addition, despite the growing use of ICT, responding using the web is still difficult for several respondents. Therefore, as detailed in the next section, it should be stressed that this mode of data collection will always be used coordinated with CATI/CAPI attempts for non-respondents.

We are planning to "push" CAWI, as the preferred option, among those who gave their e-mail address in a previous interview (about 40% of the households, currently). In addition, to a functional role in confidentiality assurance, the availability of e-mail address is also a proxy for ICT usage. However, in order to minimize non-response and have higher fieldwork efficiency, we are considering adding additional filtering requirements from previous information, such as respondents willingness to use CAWI, how familiar they are with PC or if any household member uses the web regularly.

For the remaining sample, initial mode assignment are planned to occur according to the usual criteria, mainly reliant on the existence of telephone contact for CATI versus CAPI.

3.1.2   Initial assignment and transitions between modes

At the beginning of the data collection period, from the 2nd interview forward, each household will be assigned to an initial mode (as described in the previous point), and at any point in time will only be under data collection by one mode (accessing a common platform, and thus avoiding duplications). However, it is our intention to provide as much flexibility as possible to respondents and allow for mode transitions throughout the fieldwork period. This includes, for instance, the possibility of using different modes for different individuals within a given household. These include planned transitions for CAWI non-respondents, as well as any other transitions requested by respondents.

Currently, data collection for the LFS only involves mode transitions (among CATI and CAPI) between quarters. While these also involve some challenges to handle with the introduction of web interviews, transitions within the same quarter pose a number of additional problems. In this regard, two main challenges arise: 1) technical and confidentiality issues (see next section for further details) regarding the possibility of continuing incomplete interviews after a mode transition (instead of having to start it again); 2) the short time frame for households initially assigned to CAWI.

While the specifics are still under discussion, households initially assigned to CAWI will transition to CATI/CAPI[1] in case of nonresponse only after a few days (around 4). This allows for the feasibility of still being able to get their answers by a different mode within, at the maximum, 2 weeks of data collection. Households that try to access the web survey after this point will be informed that they will be approached by interviewers (while maintaining the option to switch back to CAWI).

Additionally, it should be noted that in the case of continuity of answers – of incomplete surveys – between modes in the same occurrence (quarter) in the particular case of CAPI has limitations deriving from the offline nature of CAPI interviews. Given the online nature of CATI and CAWI, this possibility is easier to implement.

3.2 Confidentiality and CAWI authentication

The main issue regarding confidentiality, concerns providing CAWI access to information previously provided by respondents, and this is relevant both between and within occurrences (quarters).

In order to access web questionnaires for Statistics Portugal surveys, it is necessary for respondents to have occurrence specific household credentials (household ID + password). These identify the dwelling and must be used for all residents (shared login). As mentioned before, one crucial requirement of data collection for the LFS is that, even though the sample is of dwellings, it is necessary to list past and present residents in order to maintain temporal matching and unique ID for each resident.

While having these aspects into account several hypotheses were evaluated (such as individual logins and/or including a security question procedure) but the current plan involves two main dimensions:

1) Restrict initial CAWI to households with a previous e-mail address, and providing web credentials only through this communication channel (also used for password recovery). This way, only individuals from the previous group will be allowed to see personal information (e.g., name, sex, age or date of birth for initial characterization) of previous residents. If they are no longer residents, it will also be possible to ascertain if someone from the previous quarter still remains or if all have moved. In the latter case, CAWI may still be possible either, if requested to an interviewer or by (postal) sending of new credentials to the household (providing access to a questionnaire with no visible information from previous residents).

2) Disable CAWI visualization and editing access to parts of the questionnaire previously provided in other modes or by other household residents that have already completed their individual questionnaires.

3.3 Communication

Communication with respondents is always a fundamental part of a successful data collection management operation, and to maintain a high response rate. However, in situations of self-response, such as CAWI, it becomes vital given the need to motivate respondents into action, instead of a more passive role of initially waiting for contact from an interviewer. Bearing this in mind, we consider a multi-mode contact and response support communication strategy involving different dimensions.

---

[1] These are mainly transitions to CATI. Situations programmed to transition from CAWI non-response to CAPI should be minimal and limited to specific situations in order to allow for adequate logistic preparation of face-to-face fieldwork.

3.3.1 Advance letter/e-mail and reminders

In terms of initial communication, we currently use letters (via postal mail) to announce respondents they entered the LFS before their first personal interview (CAPI) and another one before their first interview by phone (CATI). With the introduction of CAWI this will be revised, at least for those initially assigned to this mode. In addition to using e-mail as the main communication channel for this group, the content of the communication itself will also be revised, namely promoting this way of answering and providing directions for how to easily do it.

In addition, given the short time initially possible for households to remain in CAWI, e-mail reminders will also be sent briefly after the beginning of data collection (3rd or 4th day) with use of SMS also being considered as a possibility to increase efficiency through the promotion of earlier responses and avoiding the need of mode transitions. Reminders are also expected to include some degree of differentiation. For example they will take into consideration if the web survey has been initiated and which mode is expected to be pursued in case of CAWI non-response.

3.3.2 Survey help and helpdesk

The questionnaire questions were reviewed to be simple, self-explanatory and as short as possible, although the Portuguese language does not always allows us to use short questions. Despite our efforts, we consider that some additional help or clarification is needed for some respondents. In addition to FAQs about the navigation of the survey, some questions and answers categories have additional help in two formats: i) always visible information, which provides complementary and helpful instructions (for the majority) to answer the questions, and ii) *Info+* button hiding additional information, not needed for the majority of respondents, clarifying, for example, some concepts and what to include/not-include.

As the same platform will be used for the three modes – CAPI, CATI and CAWI – all respondents will have access to the same information across modes, we aim to reduce mode effects resulting from the platform and help resources available.

In line with what has happened for other surveys involving CAWI at Statistics Portugal, an e-mail address and free phone number (9:30 a.m. – 9:30 p.m.) is made available for respondent support. An important characteristic of the telephonic line is that it is integrated with the data collection system. This allows for online visualization of what respondents are doing and testing (without actual editing or possibility of modification of answers) of web questionnaire response behaviour while talking with respondents. In addition, this feature also helps in an immediate and smooth transition to CATI, if necessary.

3.3.3 Website

The website survey is based on the model currently used in other surveys and has been under continuous improvement. Although the website is not designed in a mobile first approach (as recommended in literature), we recently made it more mobile compatible/friendly. The web page scale to the screen size, adjusting the visual and text elements for smaller screens, and the initial questions regarding household residents' list reformulation are some examples. However the problems associated with the presentation of additional information and optimization of some questions formats.

To add further improvements we are considering doing some user experience (UX) testing to the current ICT survey to gain some insight for the construction of the LFS website. The elements considered more critical are the login page, the initial group of questions before the LFS core questions (confirmation of address, list of residents and such) and the throughout navigation.

The first version of the LFS web survey is expected in the following months, thereupon it is expected some internal testing of the platform to look for errors. At this stage, additional improvements are expected.

4   Summary and final remarks

In this paper we present some of the key challenges that we face with the introduction of the CAWI as an additional data collection mode of the LFS. These challenges are a result of the characteristics of the survey – the length of questionnaire, follow-up up to 6 quarters, short period of response.

For the introduction of CAWI we have to decide the target respondents, including decision of whether we should use screening rules (and which ones). So far, we decided to restrict CAWI to e-mail providers, with this decision involving also consideration of confidentiality and mode transition concerns. Household approached via CAWI will receive the login credentials by e-mail (the preferred contact method), because we have a fix reference week and just 2 weeks to collect answers we are planning a 4 days of CAWI collection where e-mail (and possibly SMS) reminders will be used. Non respondents will mainly be approached by telephone for CATI interviews. With this short period of CAWI collection we expect to reach the majority of respondents willing to use CAWI, while leaving enough time left to recover non-respondents. In order to minimize mode effect, the questionnaire is common for all three modes (unimode design) with simplicity in mind to reduce friction for respondents. Additionally, we added to the web survey common help elements to all modes to compensate the absence of interviewers.

The solutions outlined before are not final, and we continue to debate them, looking for the right solution to achieve an adequate multi-mode organization and a good CAWI experience for users and good results for us – high response rate and assure good quality of final statistics. The ongoing discussion is necessary to stablish properly a set of strong guidance for CAWI implementation. Moreover, because the LFS is more demanding than other surveys, we except to use the experience gained in this process, as a benchmark for future household surveys in Statistics Portugal. We hope to learn from others' experiences and are open to any suggestions and possible solutions to improve CAWI implementation in the described context.