

Converting Social Survey Blaise 4 Questionnaires to Blaise 5: Creating a Blaise 5 Application and Modernizing the Labour Force Survey

Andy Watson and Steve Maurice, Office for National Statistics, United Kingdom

1. Introduction

The Office for National Statistics (ONS) Social Surveys Delivery (SSD) runs numerous longitudinal, annual, and ad hoc surveys. Most of these surveys are conducted as face-to-face interviews on Blaise 4. The decision has been made to upgrade these surveys to Blaise 5 so that future surveys can more easily incorporate mixed mode elements such as computer-assisted web interviewing (CAWI) and computer-assisted telephone interviewing (CATI).

The Blaise 5 uplift project was initiated to transform all the SSD surveys from Blaise 4 to Blaise 5 and all the associated legacy support systems.

1.1 The Survey for Living Cost and Food (LCF)

The Living Costs and Food Survey (LCF) collects information on spending patterns and the cost of living that reflects household budgets across the country. The study provides information about household spending patterns, which is used to update the contents of the consumer inflation basket of goods and services. It is also used to provide information about food consumption and nutrition. It is an important source of economic and social data for government and other research agencies.

The study is conducted throughout the year across the whole of the United Kingdom and is the most significant consumer study undertaken in the country. The results are essential for understanding society and planning to meet its needs.

Government departments use the results of the study to identify how and where they should be using public resources.

They use the information to check how different groups in the community are affected by existing policies and to inform future policy changes.

1.2 The Labour Force Survey (LFS)

The Labour Force Survey (LFS) is the largest regular social survey in the United Kingdom. Its purpose is to provide information on the UK labor market, which can then be used to develop, manage, evaluate, and report on labor market policies. It has been running since 1973.

The main estimates of employment and unemployment in the United Kingdom are taken from the LFS, and around 600 field interviewers and 200 telephone interviewers regularly work on the LFS. They survey a random sample of almost 45,000 UK households every 3 months.

2. Behind the Surveys

2.1 LCF

The LCF collects information on spending patterns and the cost of living that reflects household budgets across the country.

The study provides information that is used to update the contents of the consumer inflation basket of goods and services. It is also used to provide information about food consumption. It is an important source of economic and social data for government and other research agencies.

Currently, detailed expenditure data is collected by interviewers using Excel spreadsheets and embedding images of receipts within them. The process was developed at pace when the COVID-19 pandemic hit in March 2020. No user testing was carried out prior to delivery due to the pace of implementation needed to minimize the pause in data collection for the LCF.

2.1.1 LCF Methodology

The overall response rate for the LCF in Great Britain was 40% in the financial year ending (FYE) 2020, affected by fieldwork being paused because of the pandemic. This is a 3% decline when compared with FYE 2018 and FYE 2019. A total of 13,996 addresses were sampled for the LCF in Great Britain. Of these, 10% did not contain a private household and were therefore classified as ineligible.

Of the eligible sample, it was not possible to contact 9% of addresses; a further 43% refused to take part and 6% had another reason for nonresponse. Of the 5,072 responding households in Great Britain, 4,964 cooperated fully, meaning they completed both the interview and diary sections of the survey.

In FYE 2020, partial responses accounted for 2% of all cooperating households. Of these 108 partial responses, 107 occurred because one or more adults in the household refused to keep the diary but were happy to take part in the interview.

Interviewers recorded the main reason why people refused before or during an interview from a list of precoded answers. In FYE 2020, the two most cited reasons for refusing to take part in the survey were:

- Can't be bothered—24%, which remains the top reason cited as was true in the previous financial year.
- Temporarily too busy—16%, which is in line with the previous year.

Falling response rates are an acknowledged problem, and we have various initiatives to help tackle these.

2.1.2 LCF Collection

Data collection for the LCF is done face to face by field force interviewers in the homes of the respondents. Advanced letters are sent to all sampled respondents and a follow-up contact is made by the interviewers to arrange a convenient time for the interview. An In-house case management system, Casebook, is used to manage all the interviewer's cases every month.

The respondent is asked to keep track of all their spending over a two-week period, by either making note of what they purchase or preferably by keeping the receipts for all their purchases. When the two-week period ends, the interviewer returns to the respondent's address to collect this data.

The interviewer enters this detailed expenditure data using Excel spreadsheets and embedding images of receipts within them. The process was developed at pace when the pandemic hit in March 2020. No user testing was carried out prior to delivery due to the pace of implementation needed to minimize the pause in data collection for the LCF.

2.2 LFS

The LFS is a survey of households living at private addresses in the United Kingdom. Its purpose is to provide information on the UK labor market that can then be used to develop, manage, evaluate, and report on labor market policies.

Having originally been conducted every two years from 1973 the methodology, the frequency of collection and sample size changed over time, moving to its current state of collection by calendar

quarters plus an annual boost sample added in May 2006.

2.2.1 LFS Methodology

The LFS uses a rotational sampling design, whereby a household, once initially selected for interview, is retained in the sample for a total of five consecutive quarters.

The interviews are scheduled to take place exactly 13 weeks apart, so that the fifth interview takes place one year after the first.

We define Wave 1 to be the first quarter an address is selected, Wave 2 to be the second quarter in the selection, and so on. Therefore, Wave 5 is the last time that household will be interviewed for the main LFS.

2.2.2 LMS Collection

LFS fieldwork is carried out by the LFS interviewing force, which is composed of both face-to-face interviewers, who work from their homes, and telephone interviewers, who work in a centralized Telephone Operations Unit. Attempts were made to develop a web version of the questionnaire in Blaise 4, but the results weren't great.

3. LCF Questionnaire

The questionnaire is written in in Blaise 4.8 in a Computer-assisted personal interviewing (CAPI) format, with the interviewer leading the respondent through the questionnaire.

3.1 The Diary

The diary is a separate collection to the Main LCF questionnaire, which is collected by the interview and imputed into a Blaise 4 questionnaire by coders. The coders are asked to apply a Classification of Individual Consumption According to Purpose (COICOP) code for each item so that types of purchases can be analyzed at a granular level.

A small household section is completed first that includes First Name, Surname and Diary Type. Diary Type establishes if the household member is and adult or a youth. All household members are asked to fill in a diary but can refuse. This needs to be captured to establish household size.

The Usual purchases section records what type of items the household usually purchases, so that if any information is incomplete, an informed decision can be made to complete the coding.

For example, if the answer to "What type of sausages do you normally buy?" is Pork, then an entry of "Sausages - £2.50" is entered. They will be coded as Pork Sausages.

The diary collects 10 main areas of interest on expenditure within a household:

1. Food and drink
2. Takeaway food
3. Eating out
4. Clothing and footwear
5. Winnings
6. Home grown food
7. Trips abroad
8. Business refunds
9. Pocket money
10. All other purchases and payments

The diary is completed at the person level, meaning each person within the household (unless they have refused) has a record of their spending over the two-week period.

There is a Main Food Spender; this person is identified within the household as the person who does the “weekly” shopping for the household and usually has the most recorded items.

The main diary keeper keeps all large receipted “family” shopping as well as populating the “Usual Purchases.”

Each person enters their individual Spending Pattern information, which references whether the household member has any reasons for any unusual spending patterns. It may be that they were on holiday for the two weeks and did not do a weekly shop, etc.

Non-Receipted items are listed separately and usually cover small amounts of items that are bought on an ad hoc basis, like going to the newsagents and buying a paper and a can of coke, whereas receipted items are entered as a total amount on the receipt. For these items, only the shop name and the total amount are recorded at this point. The coders then enter each item separately applying COICOP codes as they go.

3.1.1 Existing Process—and Paper

Previously, each person kept a record of their individual spending by recording it all on a paper diary and giving this to the interviewer along with the receipts they had collected for the period.

Now, the interviewer sets up an Excel spreadsheet for each member of the household, which contains a tab for each day of the collection period and a separate “area” on that tab for each expenditure block. The interviewer then must scan the receipts to create a digital version.

The Excel spreadsheets and digital receipts are then submitted to HQ for the coders to pick up for manual entry into the Blaise 4 questionnaire.

3.1.2 The Original RB5 Redesign

The new requirement is to be able to enter the information into a Blaise 5 questionnaire so that some of the data can be transferred automatically.

It needs to be a single questionnaire but with the maneuverability of an Excel spreadsheet. It should still gather the same data as the multiple files.

To make the questionnaire easy to navigate between the numerous areas of interest, use of the “Go to Parallel” within the Resource Database was going to be employed. But this would mean a lot of parallels. A lot! These parallels would also have to be nested, creating another complication to the design process.

- 10 People – 14 Days – 10 areas of Expenditure
- $10 \times 14 \times 10 = 1,400$ Parallels!

The expenditure question blocks were all programmed and tested independently, so that multiple workstreams could be run. The main questionnaire with routing and navigation was developed separately and the question blocks would be entered later. During initial testing, the questionnaire contained arrays for 2 people, on 2 days, for 10 blocks, creating only 40 parallels.

The expenditure blocks were then added in successfully and a test questionnaire produced was uploaded to our internal hosting environment.

3.2 Problems with OG Design

Although the questionnaire was built successfully and worked as expected, as soon as the arrays were fully expanded to all 10 people for 14 days and all 1,400 parallels, the performance became a serious issue.

Even though the pages were built for the package, which took approximately 8 hours, it was observed that the loading time between pages was extremely slow—about 1.5 minutes per page. It was quickly decided that this was untenable and that a new solution had to be explored.

3.3 The New Proposed Solution

Previously, we had explored a “two-part” questionnaire but had discarded this solution as our in-house hosting system would not be able to support the movement between two questionnaires. When the issues arose, we quickly created a proof of concept for a two-part design to see if the issues with the in-house system could be resolved. With the support of our digital services team, a change to the core code in the servers’ token system allowed the movement from one URL to another using the same token, allowing us to jump from one questionnaire to another.

The “A” section of the new questionnaire would collect the household information—people in the house, usual purchases, etc.—and then from a menu page containing a list of valid members of the household, and a button that then would launch the “B” section.

The new “B” section would then hold the household members information at a “person-level” database. Each person would enter all their expenditure for all the days.

This reduces the number of parallels in each section to:

- A—10 Parallels (1 for each person)
- B—140 Parallels (14 days, each with 10 Expenditure blocks)

Although this required two questionnaires, doubling monthly resource requirements, this development greatly reduced both the compilation time and page load times at runtime.

3.4 Further Developments

During development of the expenditure block, it was decided that functioning summary pages would aid the interviewer in reviewing the potentially large amount of items at the end of each block. It was further discussed that it would also be helpful to enable them to edit those blocks from that review page.

If the rollout of the finalized tool is successful, Blaise 5 would be considered for the further development of the respondent-facing tool.

4. Labour Market Survey (LMS) Questionnaire

As with the LFS, the LMS questionnaire has already seen quite a few changes since its inception, and more are planned for the very near future.

4.1 Early LMS

Originally set up as a longitudinal test ahead of starting work on an LMS proper questionnaire, the longitudinal test questionnaire was web-only based and the large database included samples for all five waves for 13 cohorts. Cohorts were to be invited in one-week intervals, and the whole test was to run over five quarters mirroring LFS. As the ONS was planning to repurpose the census collection tool we had developed for business and social survey use, we were unable to host the questionnaire ourselves and were helped by colleagues at NISRA, who hosted the questionnaire for us.

The LMS had a login questionnaire with a “start survey” instruction at parallel end to redirect to the single test questionnaire GUID.

Because no training had been provided to Telephone Operations staff at this point, management of refusals, appointments, and other processing such as invite/reminder letters were conducted from a separate Blaise 4 Survey Enquiry Line (SEL) questionnaire, which every night just after midnight was populated with outcome codes from the Blaise 5 instrument downloaded and inserted via automated Manipula runs. All post-collection processing was run off this Blaise 4 instrument.

As news of a global pandemic broke, it was decided to increase the sample and ramp up the development of the longitudinal test, adding new cohorts each quarter, so that if the worst-case scenario happened and LFS collection had to stop, we might still be able to gather meaningful data. Over many weeks and months, and with many late nights working from home, the questionnaire was further developed, adding new question sections and increasing the sample. This proved challenging as not only were we making changes at a rapid pace, which was only possible thanks to Blaise 5, but we were also releasing new versions, which meant downloading a very large database from NISRA, moving collected data to the new structure (being careful to map old data to the new structure so that partials could still be completed), and packaging and reuploading to NISRA for them to install for us. The whole process with the increased sample could take 6 hours and was usually occurring every week or two.

4.2 Current LMS

At the start of 2021, as an office, it became clear that moving away from Blaise 5 was not going to happen anytime soon, and that the LMS collection in Blaise 5 had been a great success.

A demand was growing for a telephone mode LMS, and having recently been buoyed by the successful rollout of Blaise 5 training to Opinions and Lifestyle Survey staff, it was decided that Blaise 5 would be used for all questionnaires.

With plans to increase the Wave 1 sample to 12,000 cases when introducing CATI collection, it was clear that we couldn’t continue with the “One large database for the entire sample” approach we’d been struggling to maintain until this point. It was agreed that we would instead make a single questionnaire for each of the five new cohorts and start collection in any given week, which was later changed to be one Wave 1 questionnaire and one merged Wave 2–5 questionnaire each week. This required a new approach to the login questionnaire, which was developed to redirect to whichever GUID was listed alongside the Unique Access Code entered by the user. Security measures were taken to ensure that nobody could reverse engineer access to any of the cohort questionnaires directly by setting server-side variables and passing through checksum values.

A lot of work by David Kinnear was conducted to merge the SEL questionnaire into the main questionnaire and add a telephone layout that could easily be navigated by keyboard when run in CATI mode. The call treatment options of the SEL questionnaire were all added as parallel tabs of the new LMS questionnaire. This enabled us to have one questionnaire for both CAWI and CATI modes while also meeting the needs of the SEL.

A Data Collection Platform was created by our IT colleagues, which enabled us to install the Blaise questionnaires to their cloud servers for CATI collection and a nice front end was created for managing data and questionnaire removals. The stock Blaise 5 CATI Dashboard would be used for interviewers, interviewer managers, and SEL colleagues to log in and carry out their work after Roles and Skills had been set up. Unfortunately, we were not in a position to enable CAWI collection at this point, but it was planned that this Data Collection Platform would one day host our Blaise 5 web questionnaires.

With no CAWI hosting capacity, a further agreement was made with our colleagues at NISRA, who very kindly still host our LMS web questionnaire and have installed and hosted over 700 LMS questionnaires for us at this point—and a significant number of other surveys too!

Instead of downloads occurring every night as with the longitudinal test, the download frequency was increased to take place every hour to prevent chasing people who had just completed their questionnaire, with the full set of CAWI data being downloaded by an automated process and web completions and partials inserted into the CATI database. Every night a copy of the CATI data is output in the format requested by users along with a BDBX for post-collection processing to be run on.

A C# application “cloner” was developed to make short work of creating five, and eventually two, new questionnaires each week, which we now use for other surveys and is continually being developed. A separate app was created for users to run any post-collection processing, which is essentially a “Manipula Script runner” that accepts specific cohorts as input, enabling us to retire the Blaise 4 SEL questionnaire and post-collection processes entirely.

New versions of this questionnaire content are being continually developed as user needs change and a parallel run is currently taking place with a view to one day retire LFS once it’s confirmed that data sufficiently meet customer needs.

4.3 Future LMS

The future of the LMS has Blaise colleagues at ONS very excited currently, as it was recently proved that the Data Collection Platform is able to host web questionnaires successfully, with a recent run of the National Survey for Wales data being conducted online via the platform.

The Data Collection Platform manages the creation of Unique Access Codes and has its own user login page, meaning we’ll no longer have to maintain the Blaise 5 login questionnaire.

This means that alongside the further refinement and development of the current question set, we’ve been given the green light to host our own LMS CAWI questionnaire for the first time.

Work is currently being carried out at a rapid pace by David Kinnear to add a CAPI element for face-to-face interviewer collection, so that all three modes—CAWI, CATI, and CAPI—can be run on our first true mixed mode questionnaire, with all data being collected and managed in a single database. This is something we’ve dreamed of doing since we first saw Blaise 5 and it’s very exciting that we’re heading in that direction. We’re aiming to achieve this goal at the start of 2024, so if anyone has experience doing this themselves and has any advice or tips to share, we’d be very grateful to chat with you.

5. Summary/Reflections

Being in a current state of transitioning from Blaise 4 to Blaise 5 gives a distinctly unique viewpoint. It is a good opportunity to review what worked well in the old system and what we would like to see in the new system. What didn’t work well and should be left behind. What we look forward to using in the new system and features that we feel we would benefit greatly from, as well as those new available features that we are not so sure about yet, but in keeping an open mind, we may be convinced are an improved method or system.

6. References

Office for National Statistics. Methodology of LFS, Volume 1.
<https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/methodologies/labourforcesurveyuserguidance>

Office for National Statistics, LCF Methodology.

<https://www.ons.gov.uk/peoplepopulationandcommunity/personalandhouseholdfinances/expenditure/methodologies/livingcostsandfoodsurveytechnicalreportfinancialyearendingmarch2022/pdf>

Office for National Statistics. LCF Survey.

<https://www.ons.gov.uk/peoplepopulationandcommunity/personalandhouseholdfinances/incomeandwealth/methodologies/livingcostsandfoodsurvey>