

Transition from Blaise4 CATI to Blaise5 CATI at Statistics Finland

Pyry Keinonen, Juhani Saari and Petri Godenhjelm. Statistics Finland

1. Abstract

This paper discusses Statistics Finland's transition from Blaise4 CATI to Blaise5 CATI in production use during 2024. Blaise4 CATI has been used in limited capacity at Statistics Finland since 2017, when the office began its transition to Blaise5 and the in-house developed data collection management system (Ruuti). So far, no infrastructure supporting CATI mode has been created into the Ruuti system, and the interview workflow in household data collections has been implemented without the features offered by CATI.

Since the Ruuti system does not yet support business data collections, some business surveys, such as the *Job Vacancy survey (JVS)*, have been conducted with standalone Blaise4-CATI solution until 2024. The upgrade to Blaise5 CATI was piloted and tailored for the JVS, as it was the first survey to be used in Blaise5-CATI production due to its complexity and volume. The transition was implemented mainly as it was deemed necessary to discontinue Blaise4 data collections as the system's lifecycle had already reached its end.

Issues to be considered in the transition included changes in the workflow of interviewers and their supervisors, the change in the work environment from a network drive to a service run on a server and its effects on operating methods, user management and authorization, and the utilization of the new system features in work supervision and in improving the overall efficiency of the data collection. In addition, the transition had to consider changes from transmitting case data between modes during data collection.

Since 2024, Statistics Finland has sought to improve the job vacancy data collection process due to increasing survey costs and budgetary challenges. The agency had been experimenting with novel responsive data collection models, which seek to optimize the use of limited interviewer resources. These are based on modelling likely survey outcomes based on paradata from past data collection rounds together with sample unit characteristics. In its first adaptation, a sample unit's contacting order by an interviewer is determined by both their estimated response behavior as well as their overall contribution to the accuracy of JVS main estimate. Together with Blaise5 CATI-capabilities these responsive methods have improved the efficiency achieving approximately 20-25% cost savings in the required interview workhours.

2. Introduction

The transition from Blaise 4 to Blaise 5 CATI in Statistics Finland begun in the summer of 2023 as a part of larger EU Grant project aimed at improvements of the Finnish LMB Statistics, which aimed at better compliance with the requirements of the renewing legislation. The project included activities for improving Structure of Earnings Statistics (SES), Labour Cost Index (LCI) and Job Vacancy Statistics (JVS) in Finland.

This paper focuses on both the Blaise 5 CATI transition from a technical perspective and presents the results of the improvements made to JVS in Statistics Finland.

Note to the reader. The terms sample unit, case and target refer to the survey respondent.

3. Job Vacancy Survey

The Job Vacancies Survey (JVS) is a quarterly mixed-mode data collection operation managed by Statistics Finland, involving both telephone interviews (CATI) and web-based surveys (CAWI). Prior 2025 interviewers were used to collect requisite contact information for sample unit establishments, to which the initial survey request and subsequent reminder e-mails were sent. Sample units for which no valid contact email was found, the CATI mode was assigned as the primary data collection mode from the start of the survey. From 2025 onwards this prior search of contact information has been conducted centrally without interviewers for major employers with multiple establishments in the yearly sample.

To enhance data collection efficiency, Statistics Finland upgraded from the outdated Blaise 4.8 CATI platform to Blaise 5, one of the main goals being an improved and less burdensome management of the survey process, including optimized call order, case prioritization, and changes in the questionnaires during fieldwork. These features were intended to utilize the novel responsive data collection strategies, which sought to benefit the clear survey design features of the JVS, such as the relatively high predictability of response behavior within certain industries, and easily quantifiable relative importance of individual sample units to the overall estimate precision.

One key benefit of the upgrade is the ability to identify high-priority cases to allocate interviewer resource to them more effectively and in a timely manner as well as reducing the overall need for interviewer involvement by directing a greater proportion of respondents to complete the survey online. High priority targets are identified initially as sample units whose relative contribution to the overall accuracy of key survey estimate is the highest, typically representing industries with high variance in quarterly recruitment, higher personnel count and typically a lower response rate to web-mode contacting only. In addition to the more efficient utilization of diminishing interviewer resources, a new contacting mode was included in the survey model: namely a targeted postal advance letter which further encourages establishments to respond via the CAWI survey (when no e-mail contact information is available), further reducing interviewer workload.

4. Blaise 5 CATI transition

JVS was the last major data collection carried out with Blaise 4 at Statistics Finland with the past production solution having been managed in two separate surveys: the Blaise 4 CATI survey and the Blaise 5 WEB survey.

The system upgrade for CATI was carried out in two stages; in the first stage, the Blaise 4 CATI survey was upgraded to the Blaise 5 CATI survey, and in the second stage, the separate Blaise 5 CATI and WEB surveys were combined into one Blaise solution (multi-mode project).

The project work included the design and implementation of the new Blaise 5 CATI production model, related work processes, and end-user training.

4.1 Production and test environment

Blaise data collections which are conducted outside the in-house developed data collection management system Ruuti system have been implemented within Statistics Finland in such a way that separate application and database servers (on-premises) have been dedicated to the different survey modes. The servers used in CATI and CAPI modes are placed on the internal network environment and the servers used in WEB mode, respectively, on the so-called "DMZ", are intended to ensure data security. For production use Statistics Finland has one WEB server pair (application / SQL) and one CATI/CAPI server pair.

Similar setups are used in both test and production environments. This ensures compatibility of various services, applications and functions already in the testing phase.

4.1.1 Other related applications and services

Along with Blaise 5, there are two other applications/services linked to the JVS data collection process. First, one is Login portal, and the second one is the eMailer which is an in-house developed automated application for sending out survey requests via e-mail.

- 1) The login portal handles the authentication of web responders, transmits the necessary parameters and the respondent's information encrypted to the Blaise survey application. Statistics Finland has built its own decryption method. The method is always implemented survey-specifically after the Blaise package is installed into production. The manager responsible for data collection manages user IDs in the login portal and resets them if necessary (cases where there is reason to suspect that user IDs have fallen into the wrong hands due to staff changes or other factors).
- 2) The eMailer is an application that is used for sending mass messages for the respondents.

4.1.2 SQL-servers and other Blaise data

In Statistics Finland, SQL servers which are used for Blaise data collections are addressed for different Blaise application environments. Two separate SQL servers are used for Blaise data collections conducted outside the Ruuti system: a SQL server dedicated to WEB data collections and an internal network SQL server dedicated to CAPI/CATI data collections.

For the WEB data collection, only the survey response data is stored in the SQL server. Other data generated by Blaise is stored to the application server such as Audit Trail data (with session level only).

For the JVS data collection, three separate tables have been created on the internal SQL server: first for the CATI response data, second for managing the login credentials of the respondents' web survey (login portal uses this table) and a third one for managing the respondents' contact and response status information. The latter table is also used by the eMailer.

All other Blaise related data such as CATI data, Audit trail data and report-data is stored on the application server.

4.2 User Management and roles

The following role description focuses on presenting Statistics Finland's work roles from the perspective of Blaise data collection in a limited way.

To put it simply, in JVS data collection, the key actors are data collection managers, interview supervisors, interviewers and application experts:

Data collection managers prepare the prefill sample for Blaise before data collection, are responsible for reaching and communicating with respondents (eMailer etc.), and make changes to the sample data during data collection. These changes include contact information update runs and changes to the priority and thus the call order of the sample units (using manipula).

For CATI data collection, three user roles have been created in Blaise. These are the admin, supervisor and interviewer roles. AD groups have been created for the roles, which have been synchronized to the Blaise application server using the AD-sync functionality of the Blaise Server

Manager. The application specialist is responsible for the update process of these usernames at Statistics Finland.

Blaise work roles, i.e. supervisor and interviewer roles, serve as the main roles for most users. The interviewer role is limited to the use of the CATI survey application without access rights to Dashboard. The supervisor role includes extensive Blaise Dashboard access rights for the management and control of CATI data collection. At Statistics Finland, supervisors also conduct interviews in addition to their job description. Supervisors are also responsible for daily operations such as creating a daybatch or handling individual problematic cases and reporting.

The broadest range of authorizations has been granted to application specialists with admin rights. The main job description of application specialists is to ensure the technical smoothness of data collection both in WEB and CATI environment, to investigate any problems during collection, and to assist interviewers, supervisors and data collection managers in various work phases, for example, in the use of Blaise Dashboard, Manipula or how to work with the new Blaise 5 from interviewers' perspective in general.

4.3 Blaise5 JVS multi-mode solution

The JVS Blaise form version upgrade was implemented in spring 2024. Blaise 4 JVS contained two surveys: 1) the actual JVS survey and 2) the survey which interviewers filled while collecting requisite contact information about sample unit establishments (ECI) before the actual JVS survey.

The surveys were first converted to the Blaise 5 form using Blaise's own source converter. During the conversion, it was decided to continue with Blaise's traditional CATI concept so that the transition phase could be implemented with as few changes as possible to the Blaise surveys. It was not desired to introduce too many changes at once, so that the transition to Blaise 5 would be as fluent as possible.

The most typical change needs for the surveys were to update the content of the question text content, answer types, checks and role texts in general. In addition, specifications that were not considered necessary in the new model were removed from the survey's data model. These included, among other things, various Blaise 4 database views and filters that had been utilized in the Blaise 4. In Blaise 5, these functionalities were replaced by views and reports available via the Dashboard.

The largest changes during the transition phase were to the design and implementation of the user interface of the Dial and Appointment forms. The user interface design drew extensively on the expertise of the interviewer organization, improving the practical workflow and providing interviewers with the necessary information.

Further development was carried out for the data collection round of 2025 when it was decided to completely abandon the ECI survey and at the same time combine the JVS WEB and CATI solutions into one. The need for the ECI survey ceased with the JVS efficiency measures. The JVS multi-mode solution was implemented in March 2025. The multi-mode solution will facilitate data collection in the future and will enable the WEB respondents' response variable data to be transferred more easily to the CATI system thanks to a consistent data model. The data models of the previous CATI and WEB surveys differed so much that building programs to transfer response data between databases was considered laborious.

4.3.1 Dial, Appointment and Nonresponse

With the introduction of Blaise 5, it was decided to implement the Dial form from scratch. It was recognized from the Blaise 4 process that it would be good to display more information about target establishments in the new Dial form. In addition to various added contact information fields, useful information for the interviewer was added to the form, such as a) information about whether letters

had been sent to the respondent b) the value of the respondent's status variable c) call history information and more detailed appointment information, which are stored in Blaise's CATI database.

All editable input fields were mapped to the variables of the actual data collection form so that the information could be utilized in the subsequent stages of the process, whether the processing of the target proceeded to the actual interview or not. In addition to the target information, the Dial form view naturally also contains the action buttons required by the CATI process, which were mainly kept in accordance with the default settings of Blaise 5.

Perhaps the biggest difference to the standard process offered by the Blaise 5 examples was implemented by adding a so-called approval page to prevent possible “accidental mouse clicks”. The approval page was implemented using the resource database's custom page feature. The custom page includes confirmation text and two buttons; the buttons allow the interviewer to either cancel back to the Dial form or proceed to the next step in the process.

The implementation of the Appointment form largely followed the question-by-question appointment booking process as per Blaise's example. Initially, a newer appointment control was also tried, but it could not be made to work reliably in spring 2024. The selected appointment booking model also initially had a problem with the time and date fields due to the discrepancy in Finnish cultural settings. The problem was apparent in practice as the values entered in the time and date fields caused errors due to the incorrect format. These problems were fixed by CBS later.

The nonresponse block was included to the JVS main survey instead of being separated into its own. This was decided because the nonresponse block has historically always been included with main survey and this is still the case with other CAPI surveys in Statistics Finland. Also, nonresponse process follows the Blaise standards and mostly changes are cultural such as our own translations and definitions for nonresponse actions.

4.3.2 Specification file

When defining the content of the Blaise 5 Specification file, the corresponding parameters from the Blaise 4 JVS were largely utilized, which was used to achieve process similarity in the new system. The specifications that required the most special attention mainly concerned the field mappings of the new Dial and Appointment forms and the functions of the selected fields.

In contrast to the old process in Blaise 4, a new prefill variable was used to assign a priority to the respondents, which was implemented by setting the SubPriority function to the prefill variable via the Specification file. While creating a daybatch this value arranges the respondents to be called in the order desired by the manager of data collection, which allows for the initial emphasis on more aggressive reach out for targeted sample units.

4.4 Respondent status and data manipulation

In JVS data collection, the status variable determines the position of the respondent in the data collection. Status information includes, among other things, the following key code values (Table 2.):

Table 2. Example of Status code values

Status variable code value	Description
300	Initial code value for 'clean' cases
302	Completed in WEB
310	Code value for cases available for CATI
312	Completed in CATI

Above code values are used to determine which cases can be included in the CATI daybatch and are included in daily reports during data collection for case status tracking. The code value is also utilized

in communication directed at respondents during the collection process, for example by identifying those respondents to whom reminder messages are to be sent and those to be excluded.

Status code variable is manipulated both in the actual survey and by manipula programs. These programs are used mainly for transferring status code values between SQL tables and for creating daily reports for data collection managers. Also, for updating the respondent's contact information while the data collection is running.

4.4.1 Manipula programs

Manipula programs are used in JVS data collection to prefill database tables, transfer data between databases, update database case data during data collection, and generate text-based reports of these transfer runs.

Prefill manipulation runs are used only during the setup phase of a data collection round and these runs prefill target data into three different database tables: a database table dedicated to WEB collections, a database table dedicated to CATI collections, and a database table used by eMailer.

In the JVS data collection, answer data has not been transferred from WEB collection to CATI collection in the old Blaise 4 process. The transfer runs between database tables are implemented in such a way that data is synchronized from the WEB table to the contact information database table, from which the updated data is available to eMailer. Data is synchronized only for the status variable and contact information of the case. After this, the status of the case is updated to the CATI table in a second transfer run. The latter transfer run ensures that cases containing an excluding status code set in WEB response are excluded from CATI daybatch selection.

Transfer runs produce day files, manipula messages and a case tracking table about those cases that responded to the WEB survey. These reports allow data collection managers to monitor the progress of data collection in WEB survey, and if necessary, also at the case level.

In addition to the above-mentioned transfer runs, contact information collected by data collection managers while JVS data collection is already started are updated to the database tables as needed by a separate update run. Updated contact information is also synchronized to CATI-table for interviewers to use.

4.4 Changes in workflows

Even though initially the Blaise 5 survey and processes were kept as unchanged as possible compared to the old Blaise 4 era, the change required a lot of work in planning the workflow of different roles and learning new ways of working. As a result of the system update, the key changes resulted from the transition from the Blaise CATI application running on a network drive to a server-based solution.

From the developers' perspective, the transition clarified the work stages related to data processing and survey package installations. In the current model, corrections made during production can be implemented more agilely and the response data is securely on the SQL server instead of the network drive. In addition, Server Manager's AD-sync has simplified user management and streamlined processes, as user IDs do not actually need to be maintained in the application environment. The processing history of items and event logging in the new Dashboard have also made it easier and faster to investigate and resolve error situations.

Supervisors are also able to conveniently implement data collection activities via the Dashboard. The ability to change individual case information via the Dashboard user interface has been particularly appreciated. This feature has come to the fore in everyday error situations, where a case has had to be reprocessed and transferred back to the call queue with the changed information, for example due to an error in the interview situation or the system environment.

The most significant changes to the interviewer workflow focused on launching the survey via a browser instead of a desktop application. The solution was decided to avoid installing Blaise software on the interviewer workstations and thus possible problems caused by overlapping Blaise installations configured for Ruuti system processes. This was solved by creating URL shortcuts for interviewers to log in to the CATI test and production environments. It was also noticed in the interviewer workflow that the Blaise server timeout had to be manually set to allowable via the Dial form. This was since interviewers sometimes have to be inactive in the survey, for example when searching for replacement contact information through services outside of Blaise.

A notable change in the workflow for data collection managers has been the introduction of refresh runs during data collection and the use of Dashboard reports during CATI collection. The CATI paradata produced by the new system has also played a role in the improvements made to data collection methods.

5. Results

The introduction of the Blaise 5 CATI system has improved the management of paradata making it easier to monitor and adjust the survey process during data collection. This enables timely prioritization of call order and adjustments based on unpredictable issues that may arise, ultimately improving data quality. A model-based approach using auxiliary variables helps predict respondent behavior, allowing for more efficient allocation of interviewer resources by identifying which sample units (SUs) are likely to respond with minimal effort and which require more contact attempts.

A new strategy involves sending advance letters to approximately 55% of sample units, providing details about the survey and access to an online questionnaire. This aims to reduce interviewer workload and increase initial contact effectiveness, particularly for businesses in certain industries, such as schools and cultural establishments, where misidentification risk is low. These letters are sent to both first-time and second-time participants.

The advance letter's direct benefit is the reduction of interviewer effort, while its secondary benefit is the potential to improve cooperation rates. Data shows that some respondents in the web-first group still require a phone interview despite receiving an email invitation. Additionally, interviewers prioritize larger establishments and those with more complex structures in the contacting phase. The survey outcome analysis, based on contact mode and interviewer effort, reveals that web-first groups with advance letters have higher online response rates, but the letter's sole impact is not easily isolated. The findings suggest the letter increases awareness, potentially making initial contacts more productive.

In the development of a responsive data collection model, two key types of survey outcomes have been identified based on the required interviewer workload: 1) cases where no interviewer effort is needed because the survey response is obtained through automated emails or survey letters, and 2) cases that require multiple contact attempts and can only be completed through a phone interview. The goal is to prioritize contacting based on the likelihood of a sample unit (SU) responding with minimal effort, thereby optimizing interviewer resource allocation.

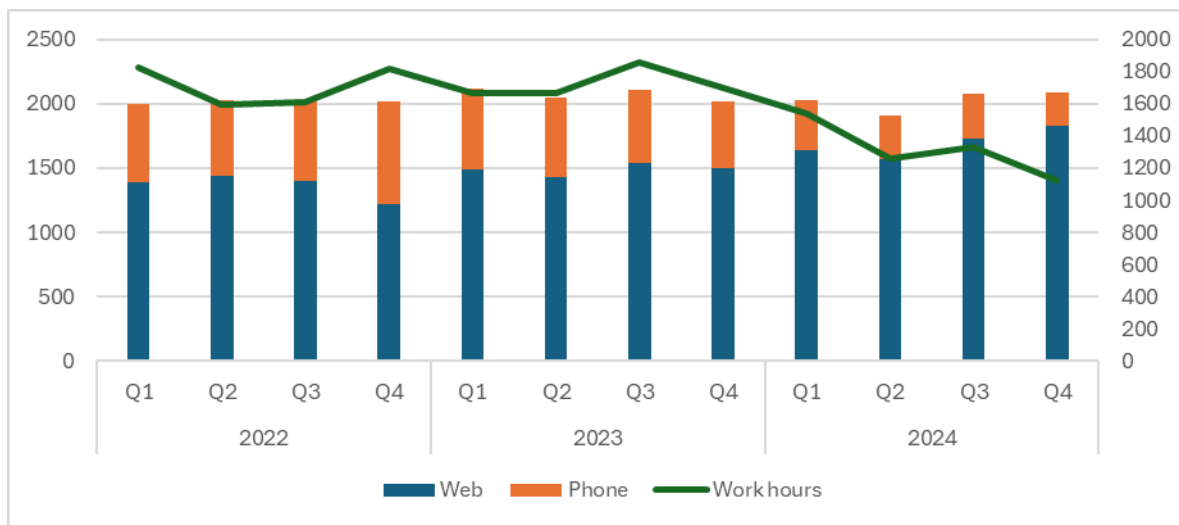
The model aims to allocate more interviewer resources to cases that need more effort, improving the efficiency of data collection. Less cooperative SUs are prioritized in the call order, allowing for a more efficient use of resources. In contrast, SUs likely to respond without an interviewer are contacted later in the process, conserving resources.

To assess the level of interviewer effort, CATI call paradata from the first quarter of data collection is used, comparing call records with online questionnaire completion times. The challenge is that call attempts are not evenly distributed, and the initial call order is influenced by the size of the

establishment. A proportional hazards survival model will be employed later to analyze the time dependency of online responses, though this is still in the early stages due to limited data. For now, logistic regression models are being used to establish initial case prioritization for the second quarter.

The first quarter of data collection between 2021 and 2024 showed significant improvements. The net response rate increased to 85.2% in Q1 2024, up from 81.7% in the previous year, while total interviewer work hours decreased by 7.9%, indicating a more efficient data collection process. Despite the improved response rate, the coefficient of variation for standard errors increased due to a more skewed variance in the sample, reflecting a decrease in the proportion of sample units reporting job vacancies. The design effect also reduced the effective sample size by 28%, affecting accuracy targets.

Figure 1. JVS response by data collection mode together with total interviewer work hours by quarter between 2022-24.



Among the key targets of the Blaise5 transition were the need to reduce survey costs and increase efficiency, while also being able to reduce the total survey error by introducing a responsive data collection model for the Finnish JVS. Second, the evaluation of the improved survey process utilizes paradata for analyzing call outcomes, interviewer work hours and respondent behavior. Based on results of the 2024 quarterly data collections, where new features were included in the survey process progressively, we were able to achieve the same comparable response rate and estimate precision (bearing in mind that part of the precision is explained by endogenous effects unrelated to the survey process).

The results of these measures are presented in figure 1, which describes the development of survey response by response mode together with total interviewer effort in terms of work hours put into contacting and interviewing sample units during the development and upgrade phase in 2024. Taken together the optimization of the data collection process has yielded a 23 percent decrease in interviewer workload while keeping the response rate at a comparable level before the cost saving measures and version upgrade took place. In these terms, the Blaise 5 upgrade and new operational procedures were indeed successful in improving the survey management and quality of the JVS.

Further improvements were noted in the allocation of interviewer resources, thanks to the new contacting mode and the enhanced Blaise 5 CATI case management system. The cooperation rate for respondents also improved, particularly in cases where advance letters were sent. These letters helped inform the designated contact persons, leading to a higher cooperation rate in several industries, particularly in sectors like agriculture, wholesale, and retail. Although the precise impact of the advance letter is difficult to isolate, it likely enhanced the effectiveness of the initial survey emails for the web-first group and contributed to the overall success of the survey process.

6. Future considerations

The development of JVS data collection in terms of Blaise solution and data collection processes will continue in 2025. It has been identified that data collection would benefit from moving response data from the WEB channel to the CATI channel. This move would better enable combined data collection for incomplete online responses. In addition, data synchronization would enable more efficient use of CATI reports, and the abandonment of old text-based reports could be considered. These changes would make it easier to organize data collection monitoring, as all relevant data collection-related reporting would be available to those who need it through the Dashboard user interface.