Statistics Netherlands

# Memorandum

# Measuring and Understanding Response Quality in the Structural Business Survey Questionnaires

**Paper prepared for the European Establishment Statistics Workshop, Nuremberg, Germany September 9-11 2013.**

Deirdre Giesen & Joep Burger[1]

August 30 2013

**summary** Response quality affects the costs for data editing and may affect the quality of the published statistics. This paper presents some first findings of a research project focusing on the quality of the raw data (as provided by respondents) of the Structural Business Survey questionnaires. Three types of quality indicators are calculated: item response, consistency and the use of "other costs". The paper explores how these indicators are related to characteristics of the respondent, the response process and the questionnaire.

**keywords** Questionnaire design, response process, perceived response burden

# 1. Introduction

Evidence-based policy making depends critically on official statistics of high quality. Quality of official statistics and how to manage it is therefore a much studied and discussed topic (e.g. Lyberg & Elvers 2003, Eurostat 2009, Snijkers & Haraldsen 2013).This paper focuses on the quality of the responses to business survey questionnaires. The goal of the research presented in this paper is 1) to develop indicators for the quality of the response to a questionnaire, that is, the "raw" data as provided by respondents, and 2) to explore how these quality indicators are related to characteristics of the responding unit (size class, type of industry), to paradata about the response process (e.g. timeliness of the response), to characteristics of the questionnaire design (e.g. mode), and, following a study from Berglund, Haraldsen and Kleven (2013), to the perceived response burden. Such indicators may be useful as they would allow a quick reaction during the data collection process. For example, in an electronic questionnaire, respondents could be warned if their response falls below a certain level of quality and be asked to change or clarify their answers. Indicators of quality of raw data are also interesting because patterns in measurement error, for example certain groups of respondents showing lower data

[1] Corresponding author Deirdre Giesen d.giesen@cbs.nl. The views expressed in this paper are those of the authors and do not necessarily reflect the policies of Statistics Netherlands

quality than others, may provide insights in where the design of the data collection can be improved.

# 2. Methods

## 2.1 Data sources

For this study on response quality we use the response to a typical official business survey, the Dutch Structural Business Survey (SBS). The SBS was chosen for several reasons: it is one of the largest business surveys conducted by Statistics Netherlands, it covers a large variety in size classes and industries, we have data on the perceived response burden of a sample of the SBS respondents, and the survey has been thoroughly redesigned.

The purpose of the redesign of the SBS was to make the questionnaires less burdensome for respondents and to reduce the costs of data collection for Statistics Netherlands. The main features of the redesign (see also Giesen & Hak 2005 and Snijkers, Onat & Vis-Visschers 2007) were: 1) the introduction of an electronic version of the questionnaire (electronic version offered only in the beginning of data collection, paper version available on request and sent with second reminder letter); 2) a reduction of the number of variables and questions; 3) shorter explanatory texts; 4) a different order of the questions: the industry specific questions were integrated with the general revenue and costs questions in the new design, whereas in the old design these were added as a separate part of the questionnaire; 5) a new layout of the paper form; and 6) a change in the sampling strategy. In the new design, businesses without employees were excluded from sample. In a previous study it was assessed that the new design of the questionnaire is associated with a lower actual and perceived response burden and a quicker response (Giesen 2013).

In the same year the new SBS was implemented, the business register of Statistics Netherlands was redesigned. This meant among others that more recent sources were used for determining the size class of businesses. We have restricted the analyses for this paper to respondents of NACE sections B, C, D and E (coded as "Manufacturing") and sections G, I, JA, JC, L, M, N and S (coded as "Commercial Services"). These sections represent about 55% of the SBS sample size. In the years 2003-2007 about 28 thousand businesses in the above mentioned NACE sections responded each year. Respectively 72% and 91 % of this selection of respondents used the electronic version of the questionnaire in 2006 and 2007.

Data on the perceived burden were obtained from a Customer Satisfaction Survey of respondents to the SBS2005 (old SBS design) and SBS2006 (new SBS design). This survey is a short telephone interview assessing the respondents' opinion about the SBS and Statistics Netherlands. Following Dale and Haraldsen (2007) the survey addressed both *perceived cognitive burden* (was answering the SBS questions easy, neither easy, nor difficult or difficult) and *perceived time burden* (was answering the questions much work, neither much, nor little work or little work). The response rate for the Customer Satisfaction Survey is typically high, around 80%. For the NACE sections analysed in this paper respectively 1262 and 1468 respondents to the SBS2005 and SBS2006 were interviewed.

## 2.2 Quality indicators

### 2.2.1 Item response

From observation of respondents in the field we know that an empty field in the SBS may mean "not applicable", "applicable but rounded off to zero" or "applicable but providing this number takes more effort than the respondent is willing or able to spend" (Giesen & Hak 2012). However, for a small set of questionnaire items we can assume that they should be applicable for all businesses: 1) the number of persons working in the business (working persons #);2) the number of full time equivalents of the persons working in the business (working persons FTE); 3) the total income; 4) the total costs; 5) the total income minus the total costs (results).For

each of these 5 variables the item response rate is seen as an indicator of the data quality. The item response rate was calculated by dividing the number of fields with zeros and the number of empty fields by the total number of fields.

### 2.2.2 Consistency

We assume that consistency between the values provided within a questionnaire makes it more plausible that the provided data are correct. Table 1 lists a set of consistency rules that are applicable for all businesses. The rule consistency rate was calculated by dividing the number of units responding consistently with the rule by the number of units eligible to the rule. Note, however, that data can be fully consistent but still be wrong, for example because the respondent reported about the wrong business unit or wrong period or just fabricated some plausible data.

**Table 1** Consistency rules as a measure of response quality.

| Rule | Description |
|---|---|
| 1a | If value for number of persons employed then value for FTE persons employed |
| 1b | If value for FTE persons employed then value for number of persons employed |
| 2 | # persons employed ≥ FTE persons employed (if value for both) |
| 3a | If value for (FTE or #) persons employed then value for costs for employees |
| 3b | If value for costs of employees then value for FTE and/or # persons employed |
| 4a | If value for # persons working then value for FTE persons working |
| 4b | If value for FTE persons working then value for # of persons working |
| 5 | # persons working ≥ FTE persons working (if value for both) |
| 6 | If value for # persons working then # persons working = # persons employed + owners/family working in business + temporary staff hired from employment agencies + other temporary staff – staff hired out |
| 7a | If value for (costs of sales + costs of employees) then value for total costs |
| 7b | If value for total costs then value for (costs of sales + costs of employees) |
| 8 | If values for costs of sales, costs of employees and total costs then (costs of sales + costs of employees) ≤ total costs |

### 2.2.3 Use of other costs / balancing item

When observing respondents complete the SBS, we often noticed that the item "other costs" was used as a balancing item to make sure that the total costs as reported in the SBS questionnaire matches the total costs as stated by the business in other reports. It is plausible that businesses report costs on the balancing item, but if its value represents a significant part of the total costs this is an indication of satisfycing. Use of the balancing item may happen more in the electronic version of the SBS, as in this version it is not possible to directly fill in a figure for the item "total costs". This is automatically calculated from all costs specifications. Respondents of the electronic questionnaire who only want to provide total costs may choose to put that number on "other costs" so that the total costs item is filled. We use the percentage of the total costs assigned to the balancing item as an indicator of data quality. This indicator is only calculated for respondents who have provided values both for the balancing item and total costs.

# 3. Results

## 3.1 Item response on core variables

Item response on the three financial variables (total income, total costs and results) was high, ranging between 94 and 99% for the total sample for the years studied. For the two variables on working persons the overall item response was clearly lower and ranged from 68% to 89%. The lower response on the personnel items may be explained by the fact that the SBS questionnaires are usually completed by respondents with access to the businesses financial data. These respondents may not always have easy access to detailed information about the

persons working in the business. The item response on the number of working persons increased after implementation of the new design, whereas the item response on the full time equivalents of the persons working decreased after implementation of the new design, especially by smaller businesses (Fig. 1). This is probably related to the fact that before the simultaneous redesign of the business register in 2006, many growing businesses were wrongfully classified as a business with less than 10 working persons. The new design did not have a structural effect on the item response on the three financial items.

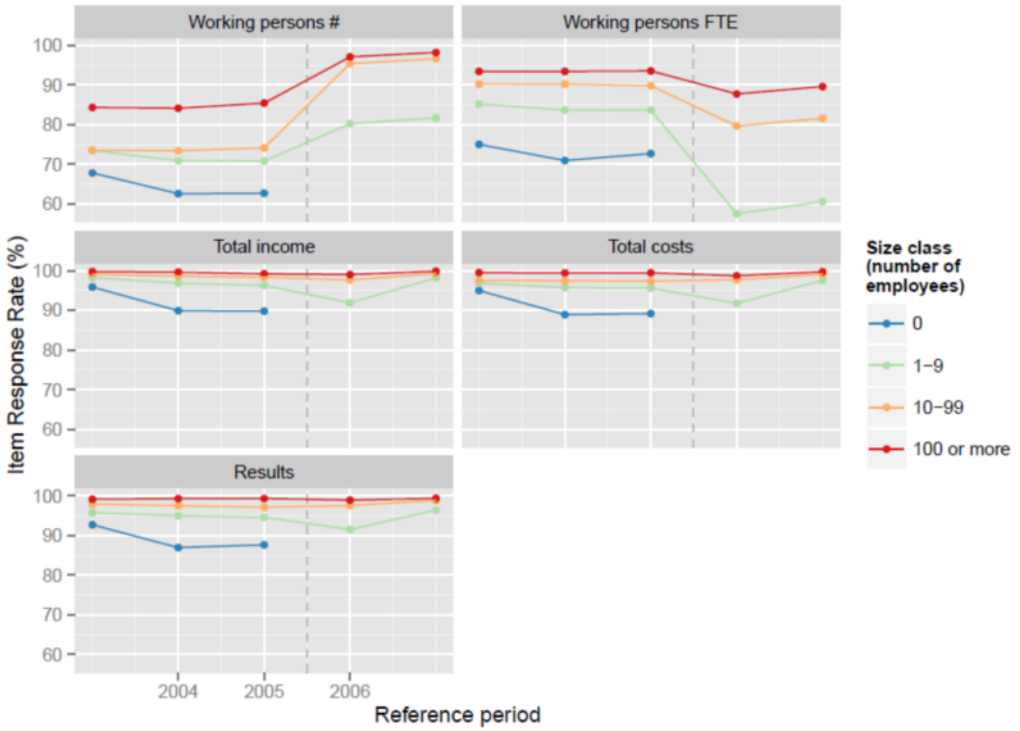


**Figure 1** Item response rate of core variables by time and size class. Dotted line denotes design change. Businesses without employees are no longer surveyed in the new design.

The changes in response rates for the two items on working persons appeared both in paper and electronic questionnaires (not shown), but may have different causes for both modes. In the paper questionnaire the main reason for the differences in item response may be due to the layout (Appendix Figs. A1 and A2). Possibly the place at the lower right end of a list of subitems conveys the message that this is an important total and makes it more likely that respondents complete this item. In the electronic questionnaire the item on number of working persons is automatically calculated from subitems on working persons (Appendix Fig. A3), whereas the item on working persons in FTE is not (Appendix Fig. A4).

The item response was always higher for the larger businesses than for the smaller businesses (Fig. 1). Businesses in Manufacturing had a slightly higher item response than businesses in Commercial Services (not shown), but most of these differences disappear when controlling for size class. Also, item response rates were often slightly higher for businesses choosing the electronic questionnaire than for businesses choosing the paper questionnaire (not shown). Item response rates did not differ between businesses responding in time, businesses responding within two months and businesses that took more than two month to respond (not shown). Item response rates were slightly higher for businesses who perceived the questionnaire as much work (as compared to those who did not find it much work) and, to a lesser extent, for respondents who perceived the questionnaires as difficult (as compared to those who did not find it difficult). These effects were, however, not consistent across size classes and branches – for some subgroups we found higher item response for respondents who reported no burden as compared to those reporting burden.

## 3.2 Consistency

Overall, the data were rather consistent with respect to the rules checked (Fig. 2). Generally between 90 to 99% of the raw responses complied with the consistency rules. With the new design, the consistency of rule 4b (if FTE than also number of persons for working persons) and rule 6 (correct calculation of number of persons working from 5 subitems) had improved from about 80% to over 95%. On the other hand, with the new design the consistency rate for rule 4a (if a figure is provided for number of working persons, working persons in FTE should also be provided) dropped steeply from 95% to 75%. The consistency rate for the other rules showed no or only very slight changes over time. For most rules larger businesses showed more consistent answers than smaller businesses (except for rules 5 and 6, which are probably easier to comply with for very small businesses that have only 1 working person). Manufacturing businesses performed better on some of the rules (especially rule 3a en rule 7b) but again these differences mostly disappeared when correcting for size class (not shown). For the rules involving accurate addition (rules 6, 7a and 8) consistency was higher and around 99% for the electronic questionnaires as expected (not shown). For rule 5 (# of persons working should be larger or equal to FTE working) paper responses performed better. This may be explained by the closer proximity of the two items on the paper form (Appendix Figs. A1 and A2) compared with the electronic form (Appendix Figs. A3 and A4). There seems to be no systematic differences in consistency for timely, late or very late respondents (not shown). As for perceived burden, we see no clear pattern for cognitive burden, but respondents who perceived the questionnaire as much work typically show a slightly higher consistency (especially for rules 3a and 7b – not shown).

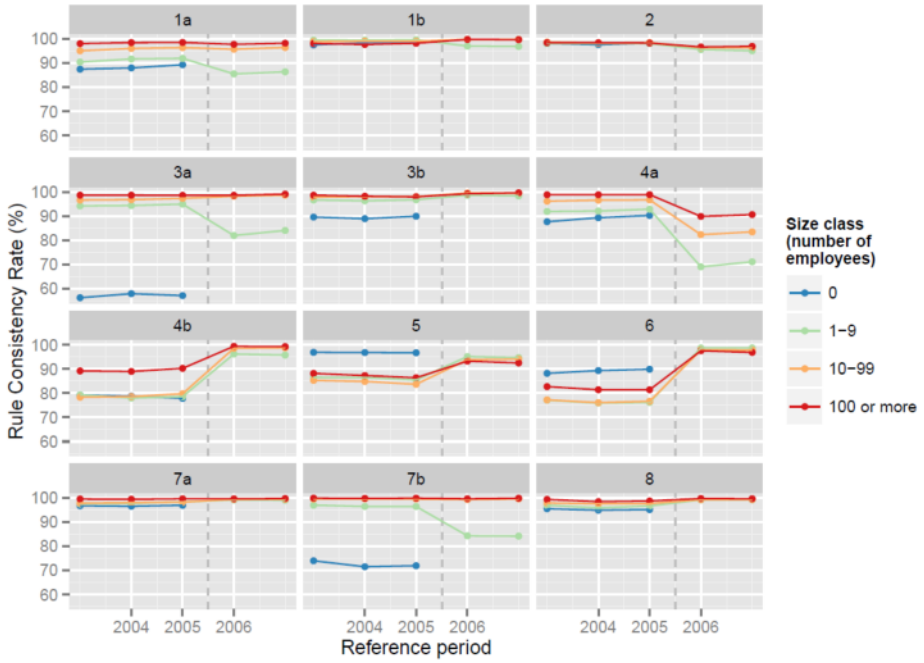


**Figure 2** Rule consistency rate by time and size class. Dotted line denotes design change. Businesses without employees are no longer surveyed in the new design.

## 3.3 Use of balancing item

Overall, about 75% of the respondents used the balancing item (left panel Fig. 3), although the value on the balancing item represented only about 3% of the total costs (right panel Fig. 3). The new design did not affect the overall use of the balancing item. Generally, smaller businesses used the balancing item less than larger businesses, but when they did they assigned a larger portion of the total costs to the balancing item. Businesses that took more than two months to respond tended to use the balancing item more often and assigned a

larger portion of the total costs to it than businesses that responded in time or within two months (not shown).As with perceived burden and item response, these patters were, however, not consistent across size classes and branches For the new design we see no differences with respect to perceived burden. With respect to the other background characteristics studied we see no clear patters in the use or relative value of the balancing item.

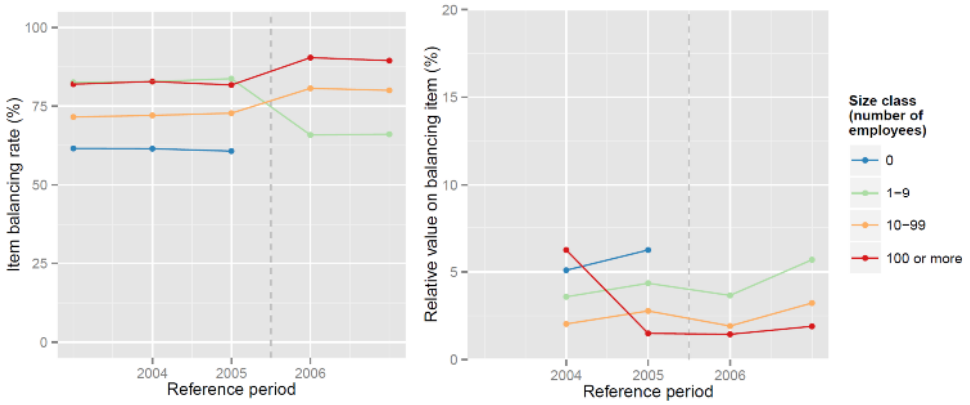


**Figure 3** Use of balancing item by time and size class. Left: percentage of businesses using the balancing item; right: percentage of total costs booked on balancing item. Dotted line denotes design change. Businesses without employees are no longer surveyed in the new design.

# 4. Discussion

The quality indicators we studied did not always show straightforward patterns related to the background variables studied. And of course, given the non-experimental nature of the data, no hard conclusions on causes and effects can be drawn from the observations made. Nevertheless, observational data can generate useful hypotheses that can drive future research. The data indicate that smaller businesses perform worse on the quality indicators studied than larger businesses. This may be due to several factors: for larger businesses respondents may have more experience with the SBS questionnaire, a better understanding of the financial terms used and access to better / more detailed bookkeeping systems than the respondents who report about smaller businesses (be it the owner who knows his business very well, but is not so familiar with bookkeeping jargon or the external bookkeeper who is familiar with the jargon but only has very limited information available about the business). Also, we see some improvement of item response and consistency due to the introduction of automatic calculation in the new survey design. Finally, we can conclude from our analyses that the new design of the SBS has improved the reporting of number of working persons and has decreased the response on the question on working persons in FTE.

The lack of systematic and large differences found may be due to the fact that our quality indicators do not discriminate enough. All in all, the quality of the indicators we explored was rather high. Possibly other indicators, with more variety, would be better detectors of patterns in response behaviour. However, it may be challenging to find quality indicators that are applicable to a wide range of questionnaires and show enough variability. On the other hand, the simple background variables we used may not be useful predictors of response quality. First of all, further analyses of our current data are needed. A first step will be to focus on one specific questionnaire and develop questionnaire specific indicators, e.g. on how many of the costs specifications are filled or industry-specific benchmark figures for plausibility. It might be interesting to explore some additional sources to find a golden standard to which we can compare the data provided by the respondent (for example the business records provided to the chamber of commerce).

# Literature

Berglund, F., Haraldsen, G. & Kleven, Ø. (2013). Causes and consequences of actual and perceived response burden based on Norwegian data. Pages 29-35 in : Giesen, D., Bavdaž, M. and Bolko, I. (Eds.)(2013). Comparative report on integration of case study results related to reduction of response burden and motivation of businesses for accurate reporting. BLUE-ETS Project Report (deliverable 8.1).

Giesen, D. (2013). Reducing Response Burden by Questionnaire design. Pages 71-77 in : Giesen, D., Bavdaž, M. and Bolko, I. (Eds.)(2013). Comparative report on integration of case study results related to reduction of response burden and motivation of businesses for accurate reporting. BLUE-ETS Project Report (deliverable 8.1).

Giesen, D & Hak, T. (2005) Revising the Structural Business Survey: From a Multi-Method Evaluation to Design. Proceedings of the Federal Committee on Statistical Methodology Research Conference, November 14-16 2005, Arlington, Virginia

Giesen, D. & Hak, T. (2012). Not Willing, Not Able. Causes of Meaurement Error in Businsess Survey. Invited paper presented at the Fourth International Conference on Establishment Surveys (ICES IV), Montreal, Canada, 11-14 juni 2012.

Lyberg, L. & Elvers, E. (2003).Survey quality issues during the last 50 years. Some observations. Proceedings of the 2003 Joint Statistical Meetings – Section on Survey Research Methods. Pp 2597-2603

Snijkers, G. & Haraldsen, G. (2013), Monitoring the Data Collection. Pp 431-457 in: Snijkers, G., H. Haraldsen, J. Jones, & D. Willimack, Designing and Conducting Business Surveys. Wiley, Hoboken.

Snijkers, G., Onat, E. & Vis-Visschers, R. (2007). The Annual Structural Business Survey: Developing and testing an electronic form. Proceedings of the Third International Conference on Establishment Surveys (ICES-3), 18-21 June, Montreal, Canada, pp 456-463. Alexandria, VA: American Statistical Association.

**Appendix** Layout of questions on working persons in paper questionnaire (old and new design) and electronic questionnaire (new design).

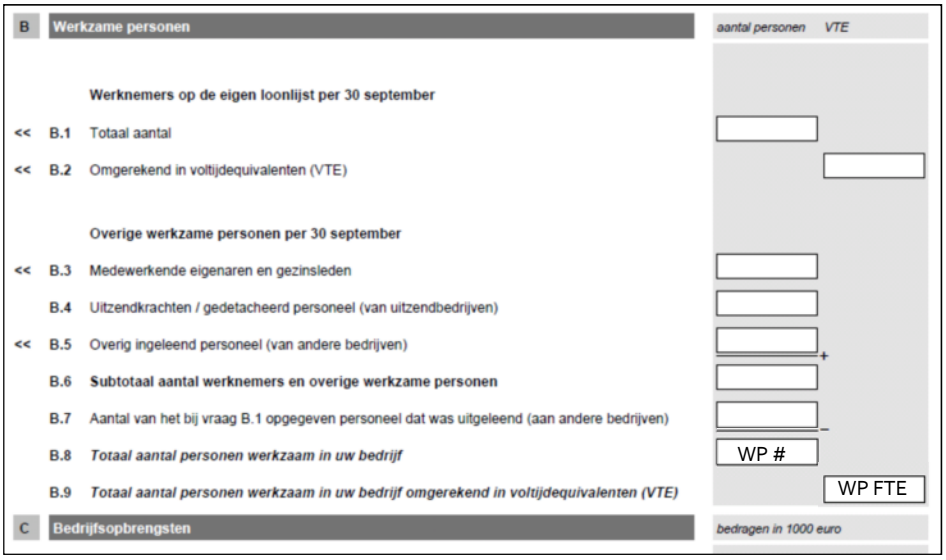


**Figure A1** Layout of question block about working persons in old design - with indication of location of question on number of working persons (WP#) and question on working persons in full time equivalents (WP FTE)

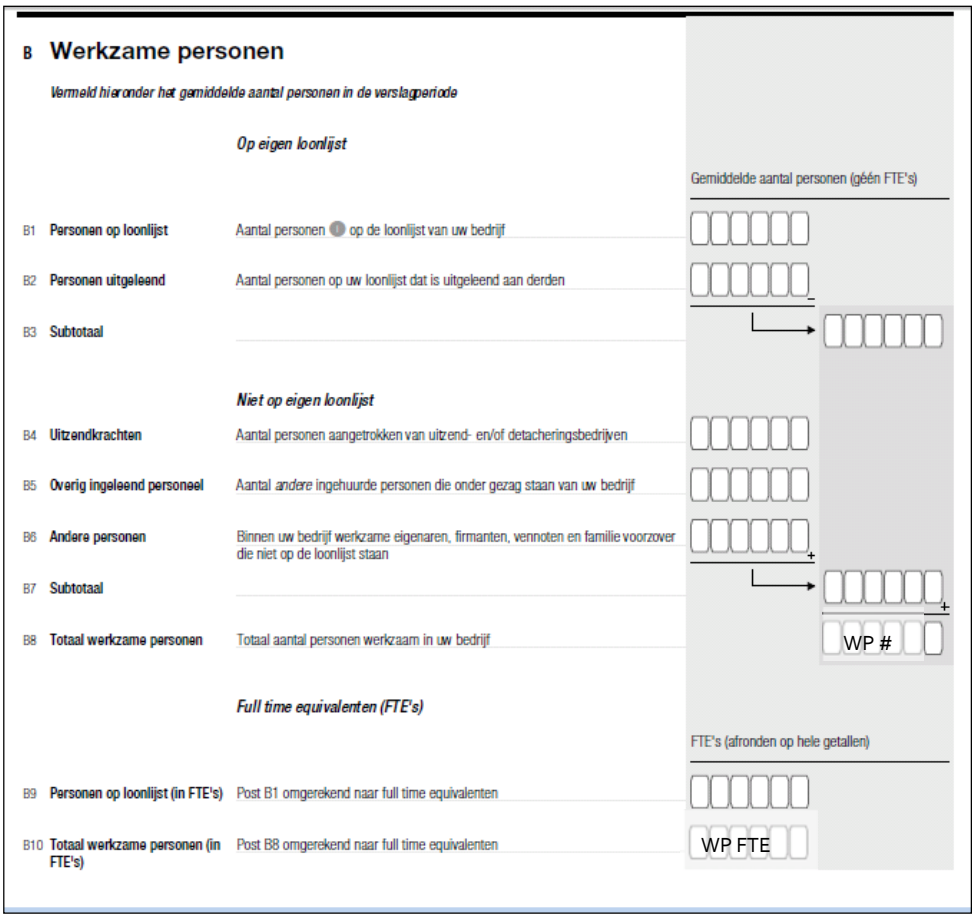


**Figure A2** Layout of question block about working persons in new design – paper version - with indication of location of question on number of working persons (WP#) and question on working persons in full time equivalents (WP FTE)

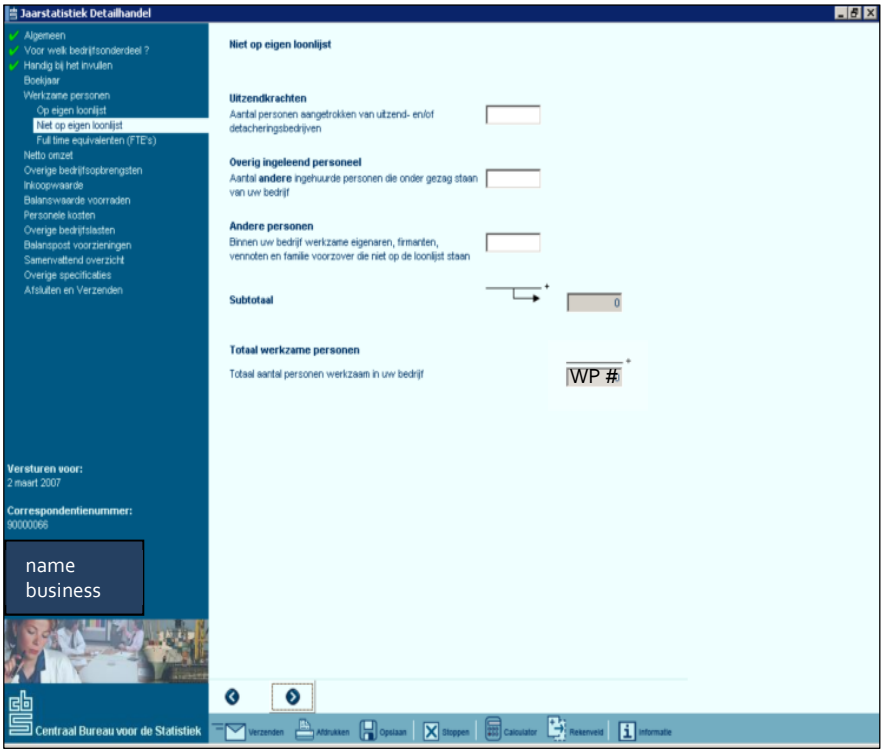


**Figure A3** Layout of question block on working persons in new design - electronic version - screen 2 of 3 screens of questions on personnel - with indication of location of question on number of working persons (WP#).

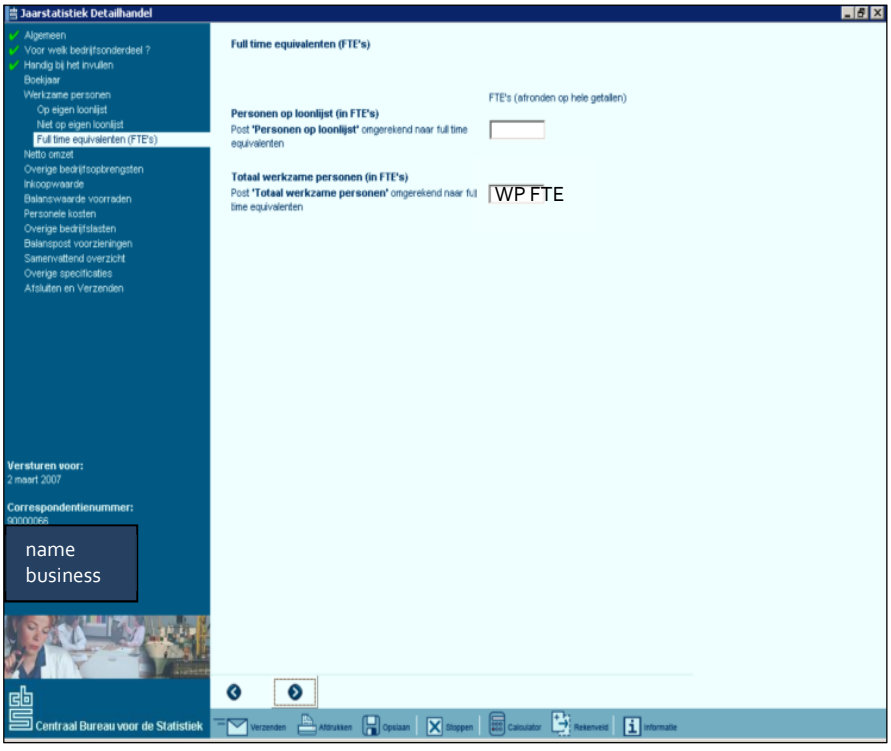


**Figure A4** Layout of question block on working persons in new design - electronic version - screen 3 of 3 screens of questions on personnel - with indication of location of question on working persons in full time equivalents (WP FTE).