

Response Quality in the Structural Business Survey Questionnaires

Paper prepared for the European Conference on Quality in Official Statistics
Vienna 2-5 June 2014

Deirdre Giesen & Joep Burger (CBS, the Netherlands)¹

The goal of the research presented in this paper is 1) to develop indicators for the quality of the response to a business survey questionnaire, 2) to explore how these quality indicators are related to characteristics the response process and the questionnaire design. For this study we use the data of the Dutch Structural Business Survey (SBS), a survey that provides input for the Dutch national accounts and the European Structural Business Statistics. The questionnaires have a core part, which is identical for all businesses and a part that varies according to size class and type of industry. In previous research we developed a basic set of quality indicators, applicable to all SBS questionnaires. The quality of these was rather high and they did not show many systematic and large differences. In this paper we explore the use of questionnaire specific indicators, specifically focusing on how operating costs are specified.

1. Introduction

The quality of the raw data as provided by respondents may affect both the costs (data editing) and the quality of the resulting statistics. The goal of the research presented in this paper is to develop indicators that can be used for the monitoring of the quality of raw data in business surveys and to explore how these indicators are related to characteristics of the response process and the questionnaire design. Detecting variations in quality and how they relate to known background characteristics can provide useful input to improving data collection strategies and tools. For example, if certain (parts of) questionnaires or certain groups of respondents perform worse than others, this information can be used to redesign the questionnaire or adapt the data collection strategy. The information may also be used to

¹ Corresponding author: 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.

provide quick feedback to respondents about the quality of their raw data. Business survey respondents who know that they provide sloppy data but think the statistical agency does not notice or does not care may have less confidence in the resulting statistics and may be less motivated to carefully complete their next questionnaires [1]. In electronic questionnaires quality indicators can be used to give direct feedback to respondents to help and motivate them to report correctly.

In the first step of this research project [2], we developed three types of quality indicators for the Dutch Structural Business Survey (SBS). This annual questionnaire collects financial data from businesses, assessing a detailed specification of the income and costs. The data are used for the European Structural Business Statistics and for the Dutch National Accounts. This survey was chosen for our research 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 can compare raw data before and after a thorough redesign of the survey (see below) and; we have data on the actual and perceived response burden of a sample of the SBS respondents.

The SBS questionnaires have a core part, which is identical for all businesses and a part that varies according to size class and type of industry. In the first step of the project we developed three types of quality indicators that could be applied to all versions of the SBS: 1) item response for five core variables that are applicable to (almost) all types of businesses (the number of persons working in the business; the number of full time equivalents of the persons working in the business, the total income; the total costs and the total income minus the total costs); 2) a set of consistency rules that are applicable for all businesses in the survey (e.g. if a respondent reports persons employed than a value for employee costs should be reported) and 3) the use of the “other costs” item (assuming that if a significant part of the total costs is put on “other” this is an indication of satisficing). These indicators were related to size class, type of industry, timeliness of the response, old versus new design of the questionnaire, mode of response (paper versus electronic) and, for a subset of respondents, also to the actual and perceived response burden.

The quality of these indicators was rather high and did not show many systematic and large differences related to the background characteristics studied. However, our analyses indicated that smaller businesses perform worse on the quality indicators than larger businesses; some

improvement in item response and consistency was seen after the introduction of the electronic questionnaire and for two specific core variables the new design seemed to have affected the item response (improving it for one variable and decreasing it for the other probably caused by the different visual presentation of these variables in the new design).

In this next step of our research we focus on one specific type of questionnaire within the SBS, so that we can explore in more detail the quality of the costs specification.

2. Methods

2.1 Data sources

We selected the SBS questionnaire that had the largest number of business units and did not involve businesses in the smallest size classes (where fewer specifications are asked). This resulted in the selection of a general Manufacturing questionnaire for businesses in size class 5 (20–49 employees in annual full time equivalents).

We used the data over the years 2003–2007. In 2003 and 2004 only paper questionnaires were available. In 2005 a pilot study was conducted, introducing an electronic version of the existing questionnaire. In 2006 a complete new design of the questionnaire was introduced, both on paper and electronically. The main features of the redesign [3][4] 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 letter²); 2) a reduction of the number of variables and questions; 3) explanatory texts shortened and the main explanatory text placed closer to the item; 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 the sample). In the same year the new design was implemented a different timing of sending out the questionnaires and reminder letters was introduced and the business register was redesigned, which meant among others that more recent sources were used for determining the size class of businesses. 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 [5].

² In 2005 due to technical problems no login codes could be provided with the second reminder letter, in 2006 and later this problem was solved.

Obviously, with multiple changes at the same time it is impossible to attribute changes in response behaviour to specific aspects of the new design.

For the questionnaire used for these analyses, the sample consisted of about 2000 business units each year, and the unit response varied between 74% to 78%. For our analyses we lost some cases when we matched the files with the raw data with files on the response process (timeliness and mode) and we dropped some cases that did not have any value for some type of income or some type of cost. This resulted in the numbers as shown in Table 1. This data set consists of 3378 unique business units. Of this group 39% were in the sample only 1 year, 26% 2 years, 20% 3 years, 10% 4 years and 5% 5 years.

Table 1 Description of data used

Year	Sample	# responses matched and selected	Response (%)	Paper (%)	Electronic (%)	In time (%)
2003	2022	1435	71	100	0	21
2004	1996	1456	73	100	0	18
2005	1958	1471	75	38	62	26
2006	1958	1427	73	27	73	34
2007	1973	1480	75	7	93	41

To study variations in data quality we compare the following

- 1) Response quality over time and especially before and after the introduction of the redesign in 2006.
- 2) Quality of the response received via paper versus electronic questionnaires.
- 3) Response quality of timely, late (up to two months) and very late (more than two months late) respondents.

2.2 *Quality indicators*

2.2.1 *Quality of calculations*

The questionnaire requires respondents to specify items and add these up in subtotals and totals. In the electronic questionnaire additions are done automatically. On paper however, errors can be made. We distinguish 6 sets:

- 1) Empty: both the subitems and the total are zero or empty. Please note that empty fields in this questionnaire can mean “not applicable”, “less than 500 euros and rounded off to zero / left empty” or “erroneously missing value”.

- 2) Unspecified: the subitems are zero or empty, but a non-zero value is reported for the total.
- 3) Underspecified: non-zero values are reported for (one or more) subitems and the total, but the actual sum of the reported subitems is smaller than the reported value of the total.
- 4) Matching: non-zero values are reported for (one or more) subitems and the total, and the actual sum of the reported subitems is equal to the reported value of the total. A match shows that the addition is done correctly and that the respondent has reported consistently for this group of items.
- 5) Over specified: non-zero values are reported for (one or more) subitems and the total, but the actual sum of the reported subitems is larger than the reported value of the total. This error can be caused by respondents who missed subitems while adding.
- 6) Imputable: non-zero values are reported for (one or more) subitems but the total is zero or empty.

One should realise that correct / consistent calculations (type 1 and type 4) are no guarantee that the provided values are valid. However, one can argue that respondents who made errors in their summations did not spend enough effort in their response task or did not have the capabilities to perform this task correctly. This may mean that it is more likely that they also made errors in other parts of the response process (e.g. looking up the correct number, making correct calculations to provide a value for a subitem).

We analysed the quality of the additions for six sets of variables for both the old and the new design. See Table 2 for an overview of these additions.

Table 2 Variables used to assess quality of calculations

Variable name	Description	Number of subitems in old design	Number of subitems in new design
SUBTOTC100000	Subtotal raw materials	2	2
SUBTOTC200000	Subtotal goods for resale	2	2
INKWRDE120000	Raw materials used	2	2
INKWRDE110000	Goods for resale sold	2	2
INKWRDE100000	Total value of raw materials used and goods sold for resale	3	4
BEDRLST310000	Total costs	12	4

2.2.2 Overall item response: fraction of subitems used to specify operating costs

For the old design the specification of operating costs consists of 70 subitems (including subtotals), for the new design this number is 55. A rough indicator of the completeness of the raw data is the fraction of subitems that are used to specify the total costs. Everything else being equal, a higher fraction of subitems indicates that the respondent has put more effort in looking up and/or calculating the subitems as required by the questionnaire.

2.2.3 Item response on core variables

For more detailed insights in the quality of the item response, we selected 11 core costs specifications, for which the definitions are comparable for both the old and the new design. These variables are listed in Table 3 below. The way items are presented (order, visual design) can differ between the old and new design, as well as the wording of the short description of the variable and the explanatory text. For the variable on wages/salaries the definition was changed in the new design (old design included subsidies and received sick pay, in the new design these were excluded). For this reason the variable name was also changed. However, this change of definition is not likely to have an impact on the item response for these variables

Table 3 Core variable used for indication of item response on costs specifications

Variable name	Description
INKWRDE120000	Raw materials used
INKWRDE110000	Goods for resale sold
INKWRDE100000	Total costs of goods manufactured and sold
LOONSOM110000/LOONSOM100002	Gross wages/salaries of employees on payroll
BEDRLST342100	Rent/lease vehicles
BEDRLST341000	Energy costs
BEDRLST343100	Rent/lease housing
BEDRLST346000	Sales costs
BEDRLST347000	Costs of communication
AFSCHRG110000	Depreciation
BEDRLST310000	Total costs

2.2.4 Fraction of total value specified

In the indicators described above we looked at the consistency of the calculations and the overall and specific item response. These indicators do not provide any information about how much of the total costs are booked on specific items. As a final indicator for this part of our project we looked at the fraction of the total value specified for the core variables listed above. In other words: how does the value specified for each of these subitems relate to the total costs specified? This relationship can expose unlikely values. Of course, for the definition of unlikely values subject matter expertise is needed.

3. Results

3.1 Quality of calculations in raw data

Figure 1 shows the type of calculations made for the six selected groups of summations. Most respondents perform the summation task correctly: Either they provide data for subitems and the totals and add up the subitems correctly (matching) or both the subitems and the totals are empty. The highest levels of consistently empty subitems and subtotals are found for the

variables SUBTOTC200000 and INKWRDE110000, which both relate to ‘goods for resale’. As this questionnaire was sent to businesses with manufacturing as main activity it seems plausible that a large proportion of these businesses do not have anything to report for ‘goods for resale’.

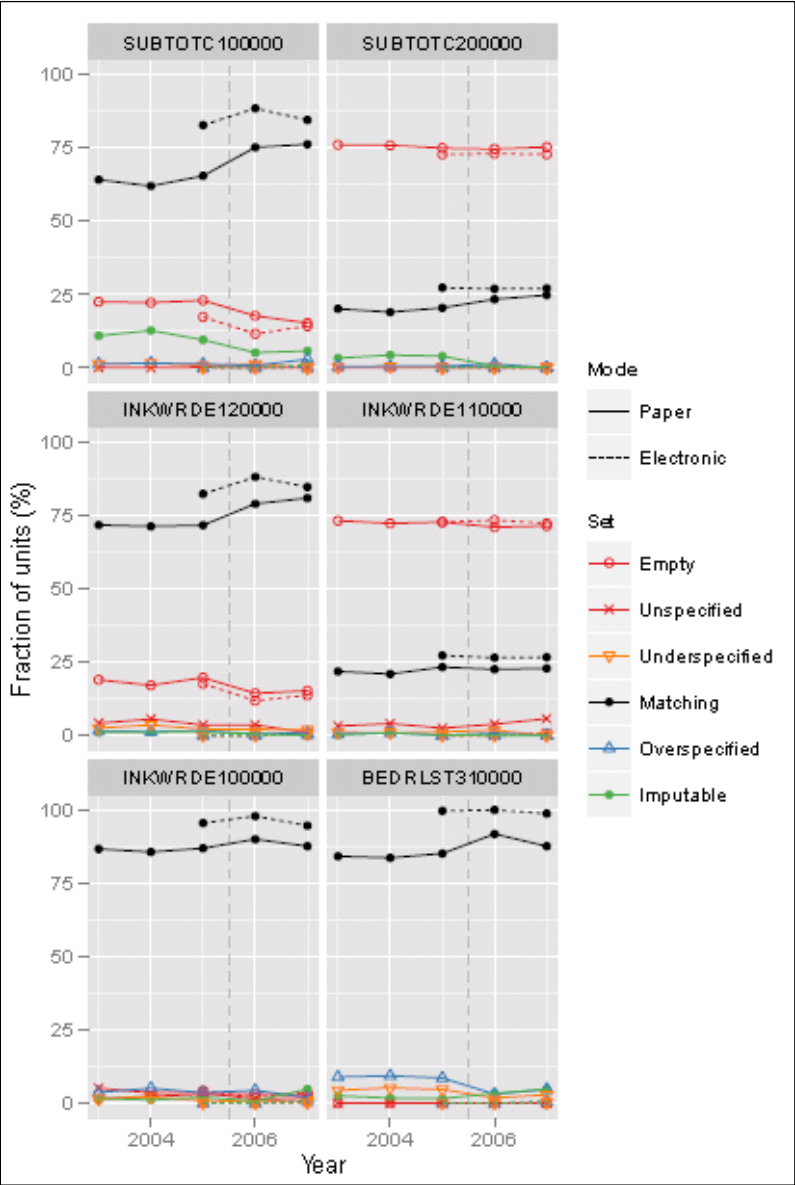


Figure 1 Correlation between the sum of the reported subitems and the reported value of the sum

Most calculation errors are made in the two groups of variables where opening stock and purchases have to be added up (SUBTOTC100000 and SUBTOTC200000) and the respondent have provided one or both of these but have not added these up themselves. This type of error has been defined as “imputable”. A closer inspection showed that of the 435 business that did not complete SUBTOTC100000 but did provide one or more subitems 41% reported both subitems (opening stock and purchases), 16% reported only the opening stock and 43% only the purchases. From the 137 business that left SUBTOTC200000 empty but provided one or

more of the related subitems, 18% has reported the two related subitems, 16% only the opening stock and 66% only the purchases. One can only speculate about the reason why SUBTOTC100000 and SUBTOTC200000 were left empty relatively often. Possibly, these respondents did not see the benefit of adding up two numbers or copying one number into a subtotal to facilitate a rather simple addition (see Appendix 1). However, this may also be an indication of erroneously missing data for some of the respondents who did not provide both subitems and an indication of some problem in the response process (difficult access to certain data and/or lack of understanding of the relevance of the requested specification).

For the total operating costs (BEDRLST310000), the most frequently made error is “overspecified”; meaning that the actual sum of the provided subitems is higher than the total provided by the respondent. In the old design this was the case for about 10 % of the questionnaires. This is consistent with the finding from previous qualitative research (an analysis of completed paper forms and field visits observing respondents completing the SBS questionnaire), which showed that especially in the old paper questionnaire respondents often forgot to include some subitems in their summations of the totals.

The new design has clearly reduced the number of calculation errors. Obviously, this is the case for the electronic questionnaires where calculations are done automatically. But interestingly, we also see an improvement in the calculations made in the paper questionnaires. This may both be due to the improved visual design of the paper questionnaire and the fact that fewer items have to be added up.

Timeliness of the response and the mode are related: later respondents are more likely to use paper, as a paper questionnaire is only provided with the second reminder letter and in the first pilot year of the electronic design no access codes to the electronic questionnaires were provided anymore with this second reminder letter. We therefore inspected if and how the quality of the calculation varies with mode and timeliness of the response. These analyses did not reveal any systematic patterns with respect to the timeliness when correcting for mode.

3.2 Overall fraction of subitems used to specify total costs

Figure 2 shows the probability density function of the fraction of subitems used to specify the total costs. For this analysis the subtotals and totals for the paper forms have been imputed to make the comparison for paper and electronic questionnaires more useful (as these are automatically calculated in the electronic questionnaire).

In the old design the total costs consisted of 70 items and in the new design – in order to reduce response burden – this number was reduced to 55. The overall fraction of subitems used is around 60%. As shown in Figure 2, it does not seem that the reduction in items in the new design has improved the response rate of the remaining items. The overall response rate is slightly higher for the electronic questionnaire. This is probably related to the fact that in the electronic form a subitem has to be filled in order to fill an automatically calculated total. Note that it was extremely rare for respondents to specify all operating costs on a single subitem (not shown). The distributions are rather stable over time. An interesting implication

is that one could use this information from $t - 1$ to assess the quality of a raw form. For example: if a respondent specifies a fraction that falls below a certain threshold a soft warning could be given in an electronic form that stimulates respondents to better specify their costs.

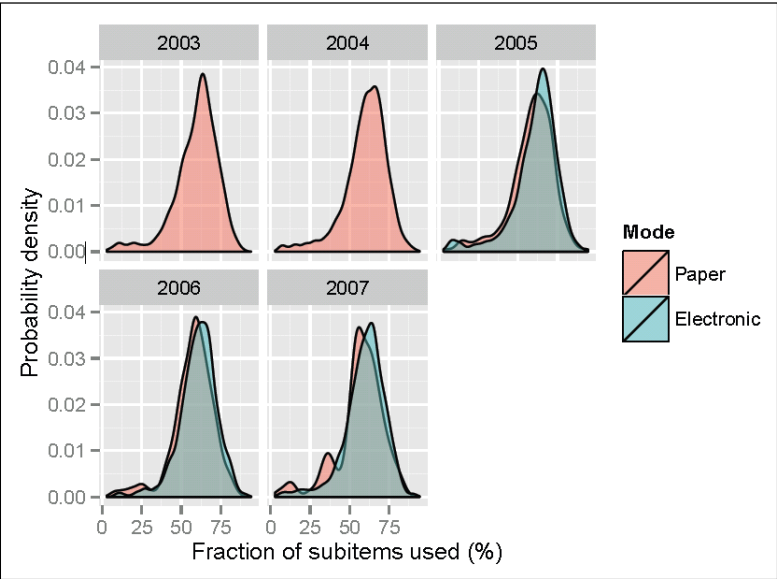


Figure 2 Probability density function of fraction of subitems used to specify operating costs

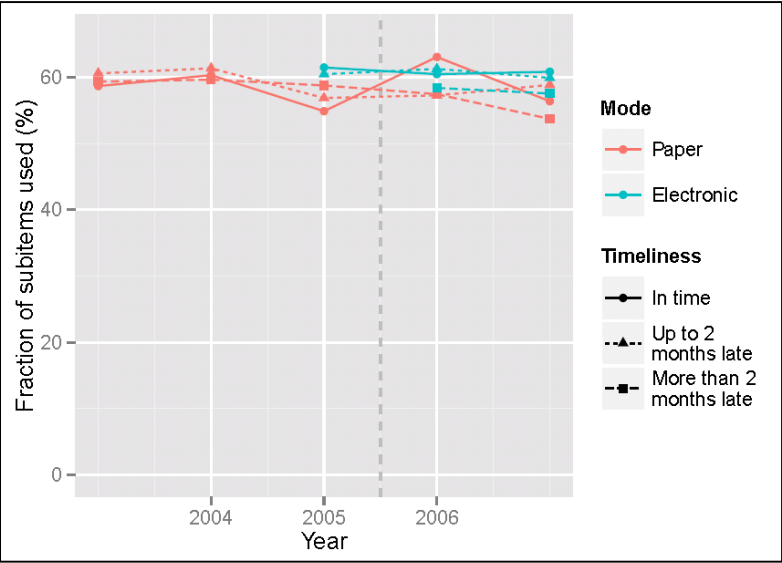


Figure 3 Fraction of subitems used by timeliness and mode of response

We assessed to what extent timely, late or very late respondents perform differently on the quality indicators. As shown in Figure 3, in the old design we see hardly any differences in the fraction specified related to timeliness for 2003 and 2004. In 2005 (old design but introduction of electronic version of old questionnaire) we see a lower item response for the timely paper respondents than for the late paper respondents. For the new design, for which also the time given to respond was reduced and the reminders sent more intensively, we see that the very late respondents (more than two months late) specify a bit less than the more

timely respondents. However in 2006 and 2007 the very late respondents perform worse than the timely respondents in both modes.

3.3 Subitem use

In Figure 4 we explore the item response on 11 core variables of the costs specifications (LOONSOM110000 and LOONSOM110002 can be treated as the same variable for this purpose). For these analyses we also imputed empty (sub) totals for the paper questionnaires where possible. Inspecting the item response over time we see differences in the overall level of item response that seem to be plausible given the content of the items. For most variables the item response stays the same but for example for variable BEDRSLT342100 we see a lower subitem use in the new design. One can only speculate that the slightly different wording, the different presentations of the explanatory text (closer to the item) and /or the different design (location on page/screen) may have affected the response process (likely the interpretation of the item). We also explored if and how timeliness of the response relates to the item response of these core variables but here also no clear patterns could be established (results not shown).

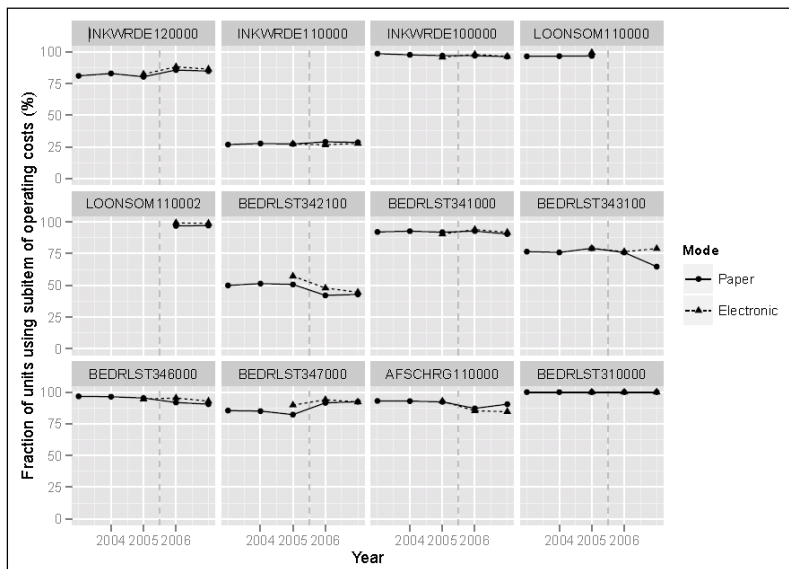


Figure 4 Item response core costs variables by time and mode

3.4 Fraction of total value specified

Figure 5 shows the distribution of what part of the total value businesses book on core subitems. As can be expected, we see that for most businesses purchases of raw materials (INKWRDE120000) and the costs for wages and salaries (LOONSOM) are a large percentage of the total costs. We see also that some business have very unlikely values, for example posts that exceed the total costs or negative values. This information could be used in electronic questionnaires to indicate possible mistakes to respondents. We see no clear relation between mode or timeliness and the relative value specified in the core subitems.

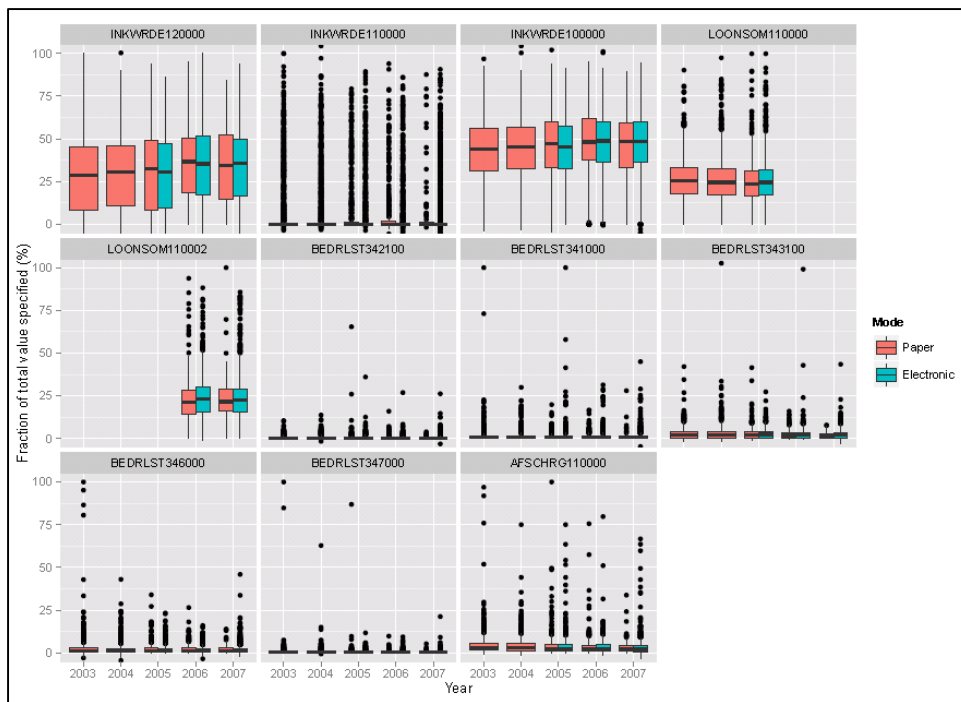


Figure 5 Fraction of total value specified for core subitems by mode of response

4. Discussion

The work presented here shows how quality indicators for raw data can be calculated for questionnaires that collect financial data by asking both totals and specifications. The methodology presented here can be applied to any type of these surveys. However, for the selection of subitems and totals and the interpretation of the results subject matter knowledge of the content of the items is essential.

The indicators we calculated can be used to provide individual feedback to respondents in order to encourage them to provide correct data. Of course, a careful balance must be sought between encouraging correct reporting and frustrating respondents with too many checks and warnings. Also, the indicators could be used to systematically monitor the quality of the raw data. Especially for indicators that have strong effects on the costs (e.g. manual editing) or quality of the resulting statistics we would like to systematically monitor when and where errors occur. This information can be used to guide the (re)design of the data collection, for example by redesigning parts of the questionnaire or the contact strategy.

As for understanding causes of data error our non-experimental data is of limited value. Differences between the old and new design cannot be attributed to specific aspects of the design. The results in Figure 4 however give a strong indication that even for “hard” financial data the wording and visual presentation of the items can affect the response process. For experiments comparing specific questionnaire characteristics the indicators presented in this paper may be useful measures.

References

- [1] Haraldsen, G. (2005). Operationalization of Response Burden through Focus Groups. In D. Hedlin, T. Dale, .Haraldsen, and J. Jones (Eds.), *Developing Methods for Assessing Perceived Response Burden*, (pp. 64-72). Research report, Stockholm: Statistics Sweden, Oslo: Statistics Norway, and London: Office for National Statistics
- [2] Giesen, D. and Burger, J. (2013). *Measuring and Understanding Response Quality in the Structural Business Survey Questionnaires*. Paper presented at the European Establishment Statistics Workshop, Nuremberg, Germany, September 9-11 2013.
- [3] 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
- [4] 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.
- [5] 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).

Appendix 1 Presentation of items involving SUBTOTC100000 and SUBTOTC200000

E Inkoopwaarde			
		Rond alle bedragen af op 1000-tallen Noteer € 23.669,- als 24	
E1	Beginvoorraad grond- en hulpstoffen Ten behoeve van het productieproces, inclusief herwaarderingen. Geen onderhanden werk	VOORRAD220100	<input type="text"/>
E2	Inkopen grond- en hulpstoffen Alle grond-, hulpstoffen en materialen die direct nodig zijn voor het productieproces ①	INKOPEN120000	<input type="text"/>
E3	Subtotaal	SUBTOTC100000	<input type="text"/> +
E4	Eindvoorraad grond- en hulpstoffen Ten behoeve van het productieproces, inclusief herwaarderingen. Geen onderhanden werk	VOORRAD220500	<input type="text"/>
E5	Inkoopwaarde grond- en hulpstoffen		<input type="text"/> -
E6	Beginvoorraad handelsgoederen Inclusief herwaarderingen	VOORRAD210100	<input type="text"/>
E7	Inkopen handelsgoederen ② Goederen die zijn ingekocht met de bedoeling deze zonder verdere bewerking door te verkopen	INKOPEN110000	<input type="text"/>
E8	Subtotaal	SUBTOTC200000	<input type="text"/> +
			INKWRDE120000 <input type="text"/>

Figure 6 Presentation of items SUBTOTC100000 in new design paper questionnaire