**Evidence-based quality improvement**

*- in questionnaire-based data collection*

Coen Hendriks, Statistics Norway, [coen.hendriks@ssb.no](mailto:coen.hendriks@ssb.no)

Dag F. Gravem, Statistics Norway, [dag.f.gravem@ssb.no](mailto:dag.f.gravem@ssb.no)

Per Ola Haugen, Statistics Norway, [per.ola.haugen@ssb.no](mailto:per.ola.haugen@ssb.no)

Katharina Rossbach, Statistics Norway, [katharina.rossbach@ssb.no](mailto:katharina.rossbach@ssb.no)

**Abstract**

*Statistics Norway has an Integrated System for Editing and Estimation (ISEE). ISEE is a generalized system for statistical data processing and the main system for data-editing in Statistics Norway. The system is metadata driven and stores data and metadata in a common information model. It is used by approximately 180 statistics. Original data from questionnaire-based data collection is stored in the system, and all edits in the data and metadata are logged. The system automatically puts a flag on each value which is being edited, either manually or automatically. The system also counts every time an automatic quality check is activated.*

*The system can be used to analyse the quality of each separate variable by counting the number of automatic and manual edits, and the number of activated quality checks. This information is collected in a quality report which is available shortly after the data collection. The variables and quality checks can easily be traced back to specific questions in the questionnaire. It can then be used to improve the questionnaire, and to evaluate the quality development after questionnaire changes or reviews. It can also be used to compare the quality of one questionnaire to another. The quality reports are also a valuable input to internal quality reviews based on the Code of Practice.*

*The paper will give a brief introduction to the system. It will give an example of how the information is being used to improve a survey questionnaire. The paper will end with some recommendations for evidence-based quality improvements in questionnaire-based data collection.*

**Keywords:** Survey data, quality improvement, editing, questionnaire design

**1. Introduction**

A condition for improved quality in the production of statistics is collaboration between the right professionals. In this case experts on IT, questionnaire design and business surveys cooperated with the quality manager. This paper will not present a new solution. We simply put together available solutions and presented them for common use in Statistics Norway. A guideline in quality management is trying to prevent errors, rather than repairing them (Haraldsen et al. 2016).

First, we give a brief introduction to the Integrated System for Editing and Estimation (ISEE). A feature of the system is the storage of information on quality. Second, we show how the quality information can be used to monitor the quality of a specific variable over time. We use the variable “cost of resold goods” from the questionnaire for the Norwegian Structural Business Survey (SBS) as an example. Third, we give some possible explanations for why the quality of this variable changed over time. Finally, we give some recommendations for further analysis.

The data collection we refer to is done using electronic questionnaires in Altinn unless other information is given. Altinn is the Norwegian platform for secure electronic data collection and exchange.

**2. A brief introduction to the Integrated System for Editing and Estimation**

ISEE is a generalized system for statistical data processing, and the main system for data editing in Statistics Norway (Haugen 2016). The system was originally developed for service production price indices back in 2005. The goal was to build a generalized and metadata driven solution for a set of approximately 20 new production price indices.

Since 2005 the system has been developed further with new functionality, improved performance and support for large volumes of data. The system is metadata driven and stores data and metadata in a common information model. It consists now of four core applications:

* *Dynarev*, for microdata editing and non-statistical edits and assignments. The most used application and the application which we refer to in rest of the paper.
* *Pris* for estimation of price indices and statistical and aggregated edits.
* *Struktur* for estimation of population aggregates and statistical and aggregated edits.
* *Driller* for drill-down, microdata editing and non-statistical edits and assign-ments. Supports large volumes of data.

Today, ISEE is used by approximately 180 statistics.

Original data from questionnaire-based data collection is stored in the system, and every edit in the data and metadata is logged. The system automatically puts a flag on each value which is being edited, either by manual or automatic editing. The system also counts every time an automatic quality check is activated.

The system can be used to analyse the quality of each separate variable by counting the number of automatic and manual edits, and the number of activated quality checks. This information is collected in quality reports which are available shortly after the data collection. The variables and quality checks can easily be traced back to specific questions in the questionnaire. It can then be used to improve the questionnaire, and to evaluate the quality development after questionnaire changes or reviews. The quality reports are also a valuable input to internal quality reviews.

ISEE has a common data model for all the applications. It has full history, dating back to 2006, for all data elements. This includes metadata, logs of the editing process and the frequency of activated quality checks. This makes the system a good source for reporting on the quality of the data.

Today, eight reports are available:

1. Share of units (questionnaires) which are edited for a series of samples
2. Number of edits per variable for a given sample
3. Number of edits per variable for a series of samples (manual and automatic)
4. Number of manual edits per variable for a series of samples
5. Number of activated quality checks by type of quality check, per sample
6. Number of submitted questionnaires per respondent, per sample
7. Completeness of the dataset. Number of missing values, in total and by variable
8. Overview of the dataset. Number of units, variables and total number of values in the dataset, by type of variable

**3. Analysing the quality of a specific variable**

A questionnaire methodologist is interested in finding out what impact changes in a questionnaire have on the quality of the data. Questionnaire changes are generally made based on qualitative findings from expert evaluation, cognitive or usability testing, but a quantification of the problems and the effects of implemented solutions is often lacking.

A reduction in the number of edits in ISEE could indicate that the questions are understood better and/or that the quality checks in the electronic questionnaire are adequate. There are many factors which can influence the quality of the data (Haraldsen et al. 2016):

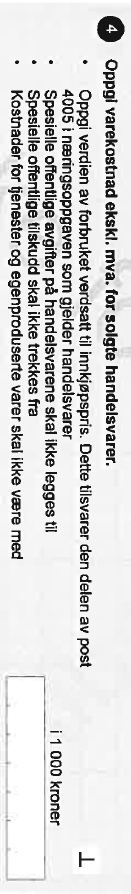
* Properties of the units in the sample
* How questions and explanations are formulated
* Visual design
* Checks and edits in the questionnaire
* Tailoring of the information flow in the questionnaire
* Other communication on the questionnaire and the survey

In the following, we will look at a specific variable from the SBS. The variable is “cost of resold goods”. We will look at how this variable was operationalized into questions for the reference years 2014-2017. From the start, the statistics divisions involved in producing structural business statistics were concerned with the quality of data produced by these questions. From a questionnaire methodology standpoint, there were issues that had been uncovered using expert appraisal. Hence, the variable was a good candidate for question redesign and pretesting.

*3.1. The reference year 2014 SBS questionnaire*

Up until 2014, the survey used Statistics Norway’s in-house developed Idun platform for business web data collection. In the Idun questionnaires, there was a single question on cost of resold goods (figure 1), which was based on the pre-existing SBS paper questionnaire. Four ballpoints offered clarifying information. The first ballpoint refers to the annual business income declaration submitted to the Tax Administration, post 4005, of which the cost of resold goods is a subtotal. The second and third ballpoint refer to various government taxes and grants that should not be added or subtracted. The fourth ballpoint refers to costs for services and self-produced goods, which is a mutually exclusive subtotal of post 4005 and should not be included.

**Figure 1. The question on cost of resold goods from the 2014 questionnaire**



One of the main problems with the question was that the term for goods for resale, *handelsvarer,* is not widely used among the respondents. One term often used is simply *varer* – goods, which also encompasses self-produced goods. As many businesses do not sell self-produced or modified goods, there is no reason for them to make any distinction. Another term used is *salgsvarer* – goods for sale – which may also be self-produced.

From a questionnaire methodology point of view, having to use lots of clarifying information on what to include and exclude is not recommended. Often, the clarifying information ought rather to be made into a series of questions instead.

*3.2. The reference year 2015 questionnaire*

Starting in 2015, the SBS questionnaire was made available through the Altinn portal, which is used for questionnaires and forms, and data exchange by all branches of government, for individuals and businesses alike. This meant a total redesign of the entire questionnaire. Additionally, paper questionnaires were no longer used.

Since finding a simple, understandable or commonly used term for *handelsvarer* seemed difficult, a different approach was tried out in the first Altinn questionnaire. Information on revenue from retail and wholesale trade, as well as from other types of industries was collected on a separate page of the questionnaire. The term *handelsvarer* was not mentioned at all on this page. Further, efforts were made to incorporate the clarifying information from the ballpoints into questions instead.

In terms of questionnaire structure, the revenue page was followed by a page containing questions on several types of costs. It was on this page that a different approach was attempted for the *handelsvarer*, by presenting the clarifying information differently and visualising a consistency check with the sales revenues reported on the previous page (figure 2).

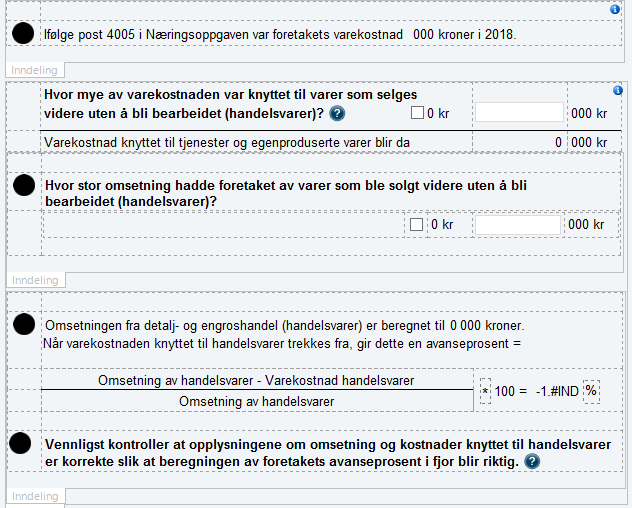
First, the sum from the business income declaration post 4005 was presented via prefill, followed by a question on how much of this sum was costs for buying goods that were “resold without further processing”. As a feature for the many businesses that do not sell goods, and to avoid empty values, a checkbox for reporting “0 kr” was added to the left of the response field.

As a way of emphasizing the distinction from costs for services and self-produced goods, this residual sum was presented as a calculation with the entered sum for costs for *handelsvarer* being subtracted from post 4005. This was all intended to convey the same meaning as the fourth ballpoint in figure 1, as well as quantifying it.

Businesses where prefill information from the business income declaration was not available got one question without reference to either the income declaration post or services and self-produced goods.

As an alternative to a hard consistency check, a suppressible (soft) check was triggered when an unlikely value was added when comparing it with sales revenues previously reported, presenting the percentage profit mark-up in another calculation. Further clarifying information was available if respondents clicked on a blue icon with a question mark – a standard feature of the Altinn platform.

**Figure 2. The question on cost of resold goods (handelsvarer) from the 2015 questionnaire in questionnaire tool InfoPath. (Dotted lines not visible for respondents.)**



*3.3. The reference year 2016 questionnaire*

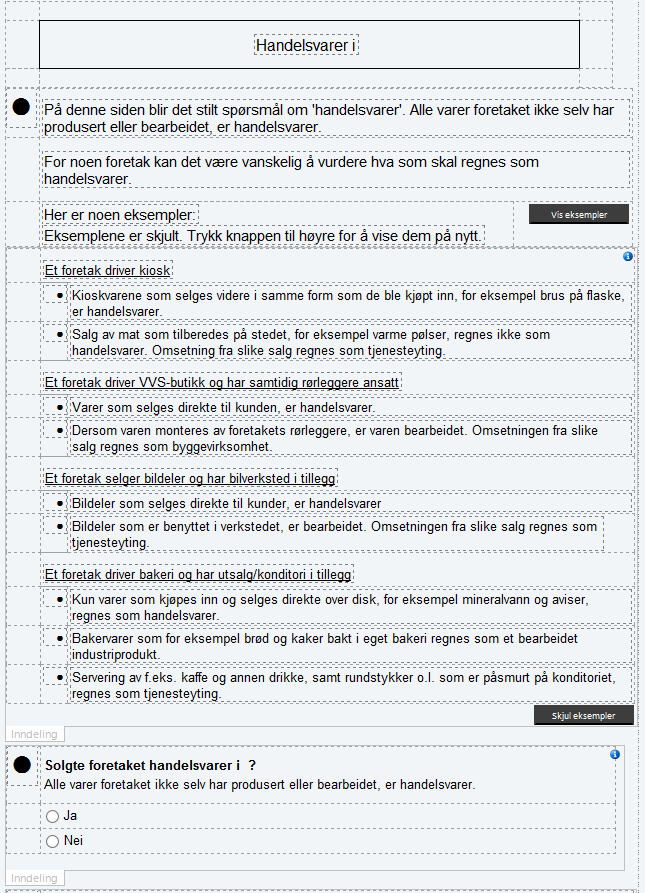
Feedback from the SBS statistics divisions indicated that the efforts put into the 2015 questions did not have the desired effect of input data quality improvement. Therefore, a series of user tests were conducted to gain a better understanding of the businesses’ problems and to come up with ideas for other approaches.

The user tests confirmed our previous knowledge of the term *handelsvarer* being difficult. They also revealed that the attempts to use an explicit description in the question (“goods that were resold without any processing”) were not intuitive. Further, the calculations and checks intended to clarify and add context were difficult to interpret and act in response to. Unfortunately, an erroneous accounting term had been used for percentage profit mark-up.

Based on this, the questions on *handelsvarer* was completely redesigned for the reference year 2016 (figure 3). To emphasize the link between revenues and costs, and to give special attention to the term and related questions, a new page for *handelsvarer* was inserted between the revenue page and the cost page. The handelsvarer page started with a definition of *handelsvarer* explaining what they are not: “All goods that the enterprise itself not has produced or processed are *handelsvarer*.”

Below the definition, examples aimed at industries where this distinction is less obvious were added: kiosks, plumbing companies, auto repair shops and pastry shops. If the enterprise had not reported wholesale or retail revenue on the previous page, this was followed by a mandatory yes/no filter question on whether the enterprise sold *handelsvarer*, with a repetition of the definition as clarifying information.

**Figure 3. The explanation on revenue from, and cost of resold goods (handelsvarer) from the 2016 questionnaire**



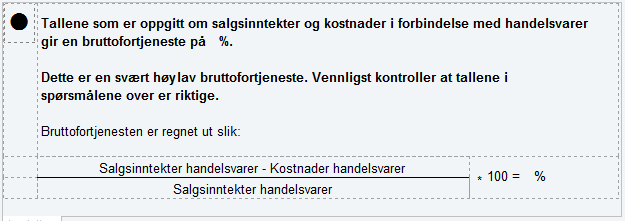
This was followed by questions on revenue from, and cost associated with *handelsvarer*, with references to the relevant business income declaration posts (figure 4).

**Figure 4. Questions on revenue from, and cost associated with handelsvarer from the 2016 questionnaire**



The consistency check was retained, this time with the correct accounting term for percentage profit mark-up (figure 5).

**Figure 5. Consistency check from the 2016 questionnaire**



Additionally, a hard consistency check was added that the costs for *handelsvarer* could not be higher than the total costs for goods:

“The costs reported for *handelsvarer* is larger than total costs for goods. Costs for *handelsvarer* is a subtotal of total costs for goods. Please correct the figures. Note that all sums must be reported in thousand kroner, e.g. a revenue of 4 914 633 shall be entered as 4 915.”

*3.4. The reference year 2017 questionnaire*

For the 2017 questionnaire, only a few modifications were made to the questionnaire. The hard consistency check of *handelsvarer* versus total costs for goods was made suppressible. In addition, two suppressible checks for cost of resold goods were incorporated in the questionnaire.

* New suppressible check 1: The suppressible check appears when the cost of resold goods equals the total cost of goods. Hence the dialog message is: “The costs reported for resold goods equals the enterprise’s total cost of goods. Only goods which are bought and resold without processing are considered resold goods. Please check the figures.”
* New suppressible check 2: This suppressible check pops up when the cost of resold goods is 80 % or more of the total cost of goods. The dialog message is as follows: “The cost of resold goods is extremely high compared to the enterprise’s total cost of goods. Only goods which are bought and resold without processing are considered resold goods. Please check the figures.”

However, a pdf instruction document was also sent as a separate message to the digital Altinn inbox of the respondent (figure 6). The respondent would find the pdf together with the link to the questionnaire. This was done at the insistence of the SBS statistics producing divisions, against the advice of data collection methodologists. The pdf instruction contained clarifying information on how to handle different key questions, and most of the space and wording concerned the interpretation of goods for resale (*handelsvarer)* for different types of businesses.

**Figure 6. Pdf instruction containing clarifying information on how to handle different key questions in the 2017 questionnaire**



**4. ISEE quality reports – what can they reveal?**

As demonstrated in the previous paragraphs, the SBS questionnaire underwent numerous and different changes between the years 2014 and 2017 on the questions on trade in goods for resale. The intention of each change was to make it easier for respondents to understand which information was required, and the changes for each reference year were user tested on the target group. The qualitative results from the user testing indicated that we were on the right track.

A decrease in the ratio of editing of the resale related cost variable could be a quantitative indication that the redesign efforts was successful. At first sight, the figures presented in table 1 show only an improvement for the variable “cost of resold goods” from the year 2016 to 2017. The ratio of manual edits on the variable increases slightly from 0.65 in 2015 to 0.67 in 2016, but decreases from 0.67 in 2016 to 0.51 in 2017.

**Table 1. Number of edits for the ISEE-variable on “cost of resold goods”, reference years 2015-2017.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Reference year** | **Data entries (original data) and non-manual changes** | **Manual edits** | **Edits in all, on the statistics variable** | **Number of reports** | **Ratio (no. of manual edits/nr. of reports)** |
| 2015 | 4857 | 2404 | 7392 | 3690 | 0,65 |
| Hereof first time registration | 4226 | 1384 | 4337 |  |  |
| Edit | 95 | 1003 | 2408 |  |  |
| Delete | 536 | 17 | 647 |  |  |
| 2016 | 3454 | 2272 | 5666 | 3369 | 0,67 |
| Hereof first time registration | 3407 | 1233 | 3410 |  |  |
| Edit | 8 | 1023 | 2215 |  |  |
| Delete | 39 | 16 | 41 |  |  |
| 2017 | 3540 | 1756 | 5254 | 3414 | 0,51 |
| Hereof first time registration | 3471 | 820 | 3471 |  |  |
| Edit | 12 | 924 | 1726 |  |  |
| Delete | 57 | 12 | 57 |  |  |

A possible explanation for not having a decrease in the ratio of manual edits from the year 2015 to 2016 might be due to a 1-year delayed effect of questionnaire improvement. Feedback from statisticians, who are responsible for this survey, indicate that sample composition might have improved and therefore helped to reduce editing. With a more detailed description for resold goods in 2016, the statisticians stated it might have affected the process of coding an enterprise’s correct industry in 2017. Thus, they could define the sample composition better for the enterprises in the survey. Hence, the redesign efforts for 2016 might have been successful, with the effect becoming visible one year later. As stated above, this is one possible explanation but there might be other reasons which we are not aware of.

The ratio of manual edits on the variable from 2016 to 2017 may suggest that the redesign efforts of the questionnaire were successful. Before jumping to any conclusions, however, it is necessary to examine other possible explanations of the observed decrease than questionnaire improvement. These include, but are not limited to:

* Learning effects among the respondents from one year to the next
* Survey communication outside of the questionnaire

It is also necessary to discuss exactly which of the changes in the questionnaire had the effect. In addition to wording and visual layout, changes in sequencing, routing and checks may have had an effect. An increase in the number and rigidness of checks could mean that the questions are in fact not easier, but rather that the editing has been outsourced to the respondents.

*4.1. Edit checks as a possible explanation*

To find out if suppressible or hard checks in the questionnaire are the reason for a reduction in editing of the variable “cost of goods”, it can be useful to have a closer look at the quality checks and the error text for each quality check. Quality checks give insight into how variables are checked internally in Statistics Norway and might indicate what is the reason behind fewer edits for a variable. For the variable “cost of resold goods” there exist three automatic quality checks (table 2).

**Table 2. Quality checks for the ISEE-variable on cost of resold goods, reference years 2016-2017.**

|  |  |  |  |
| --- | --- | --- | --- |
| **Reference year** | **Error text** | **Number of checks** | **Number of reports** |
| **2016** | “cost of resold goods” is too low | 700 | 3369 |
|  | “cost of resold goods” is negative | 83 |  |
|  | “cost of resold goods” is bigger than total cost | 2 |  |
| **2017** | “cost of resold goods” is too low (regarding total cost) | 263 | 3414 |
|  | “cost of resold goods” is negative | 136 |  |
|  | “cost of resold goods” is bigger than total cost | 1 |  |

As we can see, the number for the quality check “cost of resold goods is too low” goes down from 700 in 2016 to 263 in 2017, which might explain the decrease in editing of this variable. This quality consistency check is based both on revenue variables for resold goods and on total cost for goods. As we stated before, two suppressible checks were introduced in 2017, regarding too high values of cost of resold goods. However, the quality check that decreased in numbers is when cost of resold goods is too low. Hence, it seems as if there is no correlation between the new suppressible checks and the quality check.

It is important to note that several quality checks are based on both cost and revenue answers in one and the same quality check in ISEE. Moreover, both cost and revenue values are used to determine a consistency check (figure 5). The interaction between quality checks and how they are affected by hard and suppressible checks in Altinn, as well as the fact that both quality checks and non-suppressible/suppressible checks are based on values both cost and revenue, makes it difficult to only analyse the variable “cost of resold goods”. To sum up, the variables “cost of resold goods” and “revenue of resold goods” are strongly correlated. Therefore, it would be essential for further examination to analyse also the changes in checks - both quality as well as non-suppressible/suppressible - for the variables related to revenue of resold goods.

*4.2. Learning effects as a possible explanation*

By learning effects in a panel survey like the SBS, we mean that respondents become familiar with a question, and report more correctly from year to year. Learning effects can be related to the questionnaire/survey as a whole, or to single questions. It is fair to assume that the break from the Idun/paper 2014 questionnaire to the 2015 Altinn questionnaire was significant. The textual, visual and structural differences between the 2015 and 2016 Altinn questionnaires were also significant for the question on costs of resold goods, but from 2016 to 2017 there were only small changes to the questionnaire, and a learning effect more probable.

*4.3. External instruction as a possible explanation*

Apart from a possible learning effect, sample composition improvement and soft checks, the external instruction pdf is the only major change from 2016 to 2017. However, external/free-standing instructions and guidelines function poorly because they are often skipped by the respondents (Dillman et al 2014). Since it was sent as a separate piece of communication, could it be that it received more attention than if it had been included in the questionnaire proper?

*4.4. Possible further investigations and analysis*

Are there other changes to the questionnaire communication or editing processes that have not been properly documented, or that have escaped our attention? Regarding the possible learning effect, is it fair to assume for an annual survey? What documentation do businesses keep from year to year, and how strong is the continuity in terms of respondents and other participants in the response process? Regarding the external instruction, how are our communication efforts received and perceived?

To investigate this further, we could look at other questions, ideally with comparable initial editing rates, but where no or fewer questionnaire changes have been made, and where the 2017 external instruction would not have had a direct effect. We could also investigate whether there is a correlation between personnel continuity and data quality by checking if the contacts responsible in the businesses are the same from year to year. SBS respondents ought then to be contacted directly, using qualitative methods to gather more information. Such efforts, however, ideally should be made as soon as possible after data has been collected and ISEE data analysed.

The above descriptions and discussions illustrate the problems of using the ISEE data sources for quality assessment years after the data collection took place, with fading memories, lacking documentation and non-experimental designs. Delving more deeply into this will require considerable resources, and the best way to proceed is perhaps rather to create proper experimental designs for future SBS or other surveys. Nevertheless, the ISEE quality reports gives us a way to combine qualitative and quantitative methods for the assessment of questionnaires and response quality.

**5. Recommendations**

Quality reports from ISEE, combined with data from user test and analysis of the questionnaire, provide a good basis for evidence-based quality improvements. Using the quality reports proactively before making questionnaire changes rather than reactively, also by using embedded experiments, would likely enhance the value of the quality reports further. But they do not give information on edits and activated suppressible/non-suppressible checks which occur when the questionnaire is being filled out by the respondent. Therefore, we need paradata from the data collection process. Snijkers et al. (2013) presents examples of such useful paradata. For the Structural Business Survey 2016, Gustav Haraldsen analysed time stamp paradata, which give information when respondents logged in and out of the questionnaire (Haraldsen 2018). Moreover, he analysed the reported time for filling out the questionnaire, the time used for collecting the information, if the respondents needed to retrieve information from more than one source, and if other people needed to help in order to fill out the information. In addition, the article also includes an analysis of the respondents’ answers to which questions were perceived as difficult and the reasons why the questionnaire was burdensome. However, the 2016 paradata file did not include information about activated suppressible/non-suppressible checks. Currently, Statistics Norway are preparing a paradata framework for the Structural Business Survey 2018, which will incorporate information about triggered checks. This will allow us to conduct more complex analyses in the future.

In this paper, we have described how quality reports from ISEE can be used to improve quality in questionnaire-based data collection. In addition, these reports can be used to identify quality issues in registers which are being used for statistical purposes. Statistics Norway has agreements with approx. 30 owners of register-based sources, covering approx. 180 data deliveries. As a part of each agreement, Statistics Norway will report on the quality of the data in the register (Hendriks 2012). ISEE-quality reports can and should be developed for register data which is being edited in ISEE.

As stated before, a condition for improved quality in the production of statistics is collaboration between the right professionals. In this case experts on IT, questionnaire design and business surveys, and statisticians cooperated with the quality manager. The same applies for the analysis and interpretation of the editing data to determine quality improvements.

**6. References**

Dillman D.A., J. D. Smyth, L.M. Christian (2014) Internet, Phone, Mail and Mixed-mode Surveys: The Tailored Design Method, J. Wiley, N.Y.

Haraldsen G., C. Hendriks and A.K. Joramo Holt, Prevent or Repair? That is the question! Towards a paradigm shift in Computerized Business Web Questionnaires. Paper presented at the Nordic Statistical Meeting 2016.

Haraldsen, G. (2019), Response processes and response quality in business surveys. In: Lorenc, B. et al., ed. Cambridge Scholars Publishing, pp.155-175.

Haugen P.O. (2016), ISEE – Integrated System for Editing and Estimation. Paper presented at the Nordic Statistical Meeting 2016.

Hendriks C. (2012), Input data quality in register-based statistics – The Norwegian experience. Paper presented at the Joint Statistical Meetings 2012.

Snijkers G., G. Haraldsen, J. Jones and D.K. Willimack (2013), Designing and conducting business surveys, J.Wiley, N.Y.