Federal Institute for Vocational Education and Training

BIBB

Researching
Advising
Shaping the future

Development of a methodology for a long term strategy on the Continuing Vocational Training Survey (CVTS) CVTS3 M

Work-package 3: Modular European outline questionnaire including the description of the different modules

Work-package 4: Improvement of the operational concepts and definitions for the implementation of the survey

Paper 7: Staging/modularisation approach[1]

Bundesinstitut für Berufsbildung (BIBB)

in cooperation with

Statistics Sweden
FÁS Training and Employment Authority
3s Research Laboratory
Statistics Finland

31 May 2005

[1] Available on Circa site as Eurostat Doc. ESTAT-D5-2005-CVTS3-TF2-20-EN

1. Introduction

In this paper the advantages and the disadvantages of the implementation of a staging of the survey and modularisation of the questionnaire will be discussed and a recommendation for the CVTS3 procedure is made. In order to write this paper BIBB asked **Statistics Austria** for contributing a special expertise on staging and modularisation (see annex 1). Statistics Austria points out that a staging/modularisation approach was successfully realised in CVTS2 in Austria. **Statistics Sweden,** partner in the CVTS3-M project contributed to this work-package with an expertise about the pros and cons of staging (see annex 2). Statistics Sweden mentions especially the methodological problems using a staging/modularisation approach, and underlines that advantages of staging are contingent on method of data collection. In a second paper by **Statistics Sweden** the estimation of confidence intervals in the presence of unit non-response is treated in more detail (see annex 3).

2. The staging/modularisation approach

It can be argued that there was insufficient centralised control over the implementation of the survey in CVTS1 and to a lesser extent in CVTS2. It can also be argued that the optimal approach to data collection, which would include minimising the response burden on enterprises, is not necessarily guaranteed by implementing a survey with a single stage data collection and a single questionnaire applied in the different data collection methods (see WP 5). The main feature of the staging/modularisation approach is that in using some screening questions in the first stage of the survey (pre-survey) it is possible to develop tailor-made modules for the second stage of the survey (main survey). The aims of the staging of the survey and modularisation of the questionnaire are:

- To provide valid data with a high response-rate within a reasonable short period;
- To collect and check the basic structural data of enterprises (size of enterprises, economic sectors) as well as the basic data on training of enterprises;
- To structure the questions in modules in order to develop a tailor-made main questionnaire (modularisation);
- To reduce the burden on enterprises in filling in the main questionnaire;
- To fix the interview-partner of the enterprises for appropriate contacts in the course of the survey;
- To assure an optimal steering of the survey.

3. Experiences with a staging/modularisation approach in the context of CVTS (CVTS1 and CVTS2)

In CVTS1, because of differing national conditions, relatively broad guidelines were established for the data-collection approach. Several countries used some sort of a staged survey approach. In the first stage they used a short questionnaire to collect basic data. The distinction between training and non-training enterprises was measured in this way in Belgium, Denmark, Spain and Portugal. A more comprehensive and differentiated data-collection approach was used in the Netherlands and Germany: these countries reported at that time a successful implementation of a staged survey combined with a modularisation of the questionnaire.

For CVTS2 the working group preparing the survey advised against a staging/modularisation approach. Some countries using a postal survey without staging/modularisation achieved very low unit response rates (Denmark, Germany), whereas others like Portugal had higher

response rates. Spain, France and the Netherlands used in CVTS2 a staging approach for screening the training and non-training enterprises. Only Austria conducted a staged survey using a pre-survey and main-survey and a modularised questionnaire (see annex 1), achieving a response rate in the medium range. Even though a possible effect of using a staging approach on the response rate is not clear cut, the implementation of CVTS2 in Austria clearly had advantages in the given situation. The reported reason to use the staging/modularisation approach was to reach a reasonable response rate within a very short period. Although Austria implemented the survey much later than the other countries, the complete survey, starting with the fieldwork up to finishing the final data set, was accomplished within eight months.

Five countries (Belgium, Denmark, Germany, Italy and Portugal) had implemented a staged survey in the first survey, but not in the second survey. The response rates decreased considerably in some of these countries. In contrast, the response rates increased in Spain, France and the Netherlands, countries in which both surveys (CVTS1 as well as CVTS2) were staged. At the time being this may not be regarded as an effect of staging, as detailed analysis would be required, including other factors impacting on response rates. In WP5 we will give special consideration to methods of data collection, the application of modular questionnaires and staging of the survey.

4. Principal considerations and critical remarks using the staging/modularisation approach

In the following we will discuss some principal considerations using the staging/modularisation approach - some have already been mentioned before in the methods report - and we will cite some of the critical remarks forwarded in the reports of Statistics Austria and Sweden (see annexes). In order to be able to discuss the staging/modularisation approach we systematize the pro and con arguments:

Pre-survey: All enterprises

Function of the pre-survey

The aim of the pre-survey is to collect data for the preparation of the modules used later. The first aim is to separate enterprises into training and non-training enterprises. The second aim is to prepare the modules for the main survey, which will be used for questioning the training enterprises.

Mode of data-collection of the pre-survey

The pre-survey can be a short postal questionnaire or a telephone interview. Statistics Austria used a combination of postal questionnaires and intensive telephone contacts with the key contact person in the enterprise. The first stage questions should be in such a form that all questions could be answered without any time-consuming in-house inquiries, and thus response could be very fast and unproblematic. However, Statistics Sweden points out here that a good pre-survey will take a rather long period for the data collection and will be costly.

Content of the pre-survey

The pre-survey should only include structural questions (e.g. size of enterprises, economic sector, number of local units) and questions relating to the existence of any continuing

vocational training (e.g. provision of training courses, division of training courses into internal/external courses, provision of "other forms" of training) or initial vocational training. However, the definition of a "training enterprise" seems still to be debatable because of problems in identifying the "other forms" of continuing vocational training. Statistics Austria reported problems with the collection of data for the "other forms" of continuing vocational training. Instead of asking for the occurrence of the "other forms" of continuing vocational training separately, the measures in this field were summed up. Thus the screening question in this form in Austria caused an underestimation of the "other forms" of continuing vocational training. Statistics Sweden suggests that in order to produce useful results questions about the participation of employees should be included in the pre-survey. In the pre-surveys used until now, such questions asking the detailed the number of participants were not used.

Statistical questions relating to pre-survey

Statistics Sweden notes that it is of vital importance to get a very high response rate in the first stage. If the response rate of the pre-survey is low, assumptions, especially concerning the weighting, are necessary.

Key contact person in the enterprise

The staging/modularisation approach makes it possible to identify the key contact person in the enterprise. The name, telephone-number, e-mail address of a key-contact person being responsible for continuing vocational training and initial vocational training (if different) within the enterprise is known. The key contact will be the contact person for further inquiries during the main survey period.

Use of the pre-survey questionnaire within the enterprise

The staging/modularisation approach will enable the key contact person of the enterprise to find the persons who can answer the main questions best. Only in small enterprises the questionnaire will be answered by one single person (e.g. the owner / director). In bigger enterprises, in addition to the management, the person(s) responsible for training and the finance department can be competent informants for some of the questions. Especially Statistics Austria mentions here, that a potential risk is that the questionnaire might "go astray" within the enterprises. The staging/modularisation approach could minimize this risk. Statistics Sweden mentions a problem when informing the enterprises about the pre-survey and the follow-up main survey. Enterprises could decide not to participate in the survey at all.

Dissemination of pre-survey results

The conduct of a pre-survey makes it possible to collect first data for training as well as non-training enterprises. First key results on the supply of continuing vocational training in training enterprises and data on the reasons not to train of non-training enterprises could be available. These basic data could be published rather quickly. In CVTS1 Germany produced a short brochure, which was sent to the enterprises of the sample.

Main survey: Non-training enterprises and non-response enterprises

Modules for non-training enterprises

The staging/modularisation approach makes it possible to divide enterprises into training and non-training enterprises and to send out tailor made questionnaires in the second phase. For the non-training enterprises, two modules could be developed. First of all a module for those enterprises which do not train in the reference year, but trained in the years before or are planning to train in the following years. We call these non-training enterprises "casual" non-training enterprises. Secondly, we can split a module for the non-training enterprises, which do not train at all in a given period (say 6 years) - the so-called "notorious" non-training enterprises. Data concerning non-training enterprises could be collected by a short telephone interview. Statistics Sweden states that the staging/modularisation approach may be a clear alternative for a country with a high ratio of non-training enterprises. Using this approach in these countries makes it possible to get higher response rates and better information for non-training enterprises. Statistics Sweden argues, however, that there will be a risk of lower response rates for training enterprises.

Checking of non-response enterprises

The staging/modularisation approach may also allow for better contacts with the non-response enterprises. Checking the reasons for unit non-response by a short telephone interview is not only from a methodological viewpoint important. Experiences with CVTS1 and CVTS2 show that often enterprises (especially smaller enterprises) do not think that they could provide relevant information for the survey. However, in reality they can. Statistics Austria thinks this is a main advantage of the staging/modularisation approach. They stipulate that a unit non-response analysis can enhance the quality of the data decisively.

Main survey: Training enterprises

Function of the main survey (modularisation)

The staging of the survey, the tailor-made instrument, would result in different versions of the main questionnaire (modularisation). According to the information gathered in the first stage and based on a typology of enterprises it is possible to customise the questionnaire according to specific training activities.

Mode of data-collection of the main -survey

For the main survey different data-collection methods can be used: postal survey, telephone interviews with or without the use of CATI, face-to-face interviews with or without the use of CAPI, providing the information by e-mail. Selection of these methods might be related to the particular responses to the pre-survey. Statistics Sweden points out that a staging approach is mainly an option to consider when using a postal questionnaire.

Relation between pre-survey/main survey (content of questions)

Statistics Sweden points out that a staging/modularisation approach is dependent on correct answers in the pre-survey, otherwise a correct modularisation will not be possible. It is more difficult in a staging/modularisation approach to check the correctness of answers than it is the case in a one-stage survey, as less information on the responding enterprise is available.

Relation between pre-survey/main survey (response rates)

Statistics Sweden notes that is of vital importance to reach high unit response rates in the pre-survey (at least 70%) in order to realise a good sample for the main survey. Experiences with the staging/modularisation approach show that it is possible to reach a high response rate with a pre-survey. The response rate of CVTS1 in Germany was around 60 %. Statistics Austria reports for the CVTS2 pre-survey an average response rate of 78 % or 90 %, depending on the intensity with which enterprises were contacted.

Relation between pre-survey/main survey (weighting procedure)

The interpretation of the relation between the pre-survey and the main-survey with respect to the weighting procedure used and the consequences for the staging/modularisation approach is seen differently between Austria and Sweden. Statistics Sweden mentions that a staging/modularisation approach according to the statistical theory implies a two-phase sample design. According to Statistics Sweden, this requires the application of other statistical theories than in the case of a one-stage survey. This especially affects the use of weighting procedures. Variance in results will be higher in a two-phase sample design than in a one-phase sample design. The effect of a two-phase sampling approach on the estimation of confidence intervals is discussed in their paper. Statistics Austria, however, argues that a staging/modularisation approach is not a two-phase sample design. From their point of view, the calculation of the variance can be based on a homogeneous sample.

Response burden

The modularisation of the main questionnaire (especially if it is a postal survey) makes it possible to ask only questions that are relevant for the enterprises concerned. The motivation of the contact person in the enterprise will be higher if he or she is not asked unsuitable or irrelevant questions and this may increase readiness to respond. Statistics Sweden underlines that by using Internet, CAPI or CATI as methods of data-collection the response burden for enterprises will also not be so high, because not-relevant questions will not be asked.

Flexibility of the survey-instrument

Staging could make it possible to administer the single modules in a flexible way and in the most appropriate form. Questions concerning the training policy, the training subjects and training providers (often causing assigning problems) could be dealt with a competent person by telephone or face-to-face.

Use of modularised main questionnaire within the enterprise

The single blocks of questions in the main survey enable an efficient steering of the questionnaire. The survey institute will receive information about missing data at an early stage and can query the enterprises in an appropriate way. Frequently the training department does not collect the quantitative data, but it can solve the missing data problems in co-operation with the departments for finance or personnel at an early stage. Moreover some of the “hard” data can be substituted by plausible estimations, which is better than using imputations afterwards.

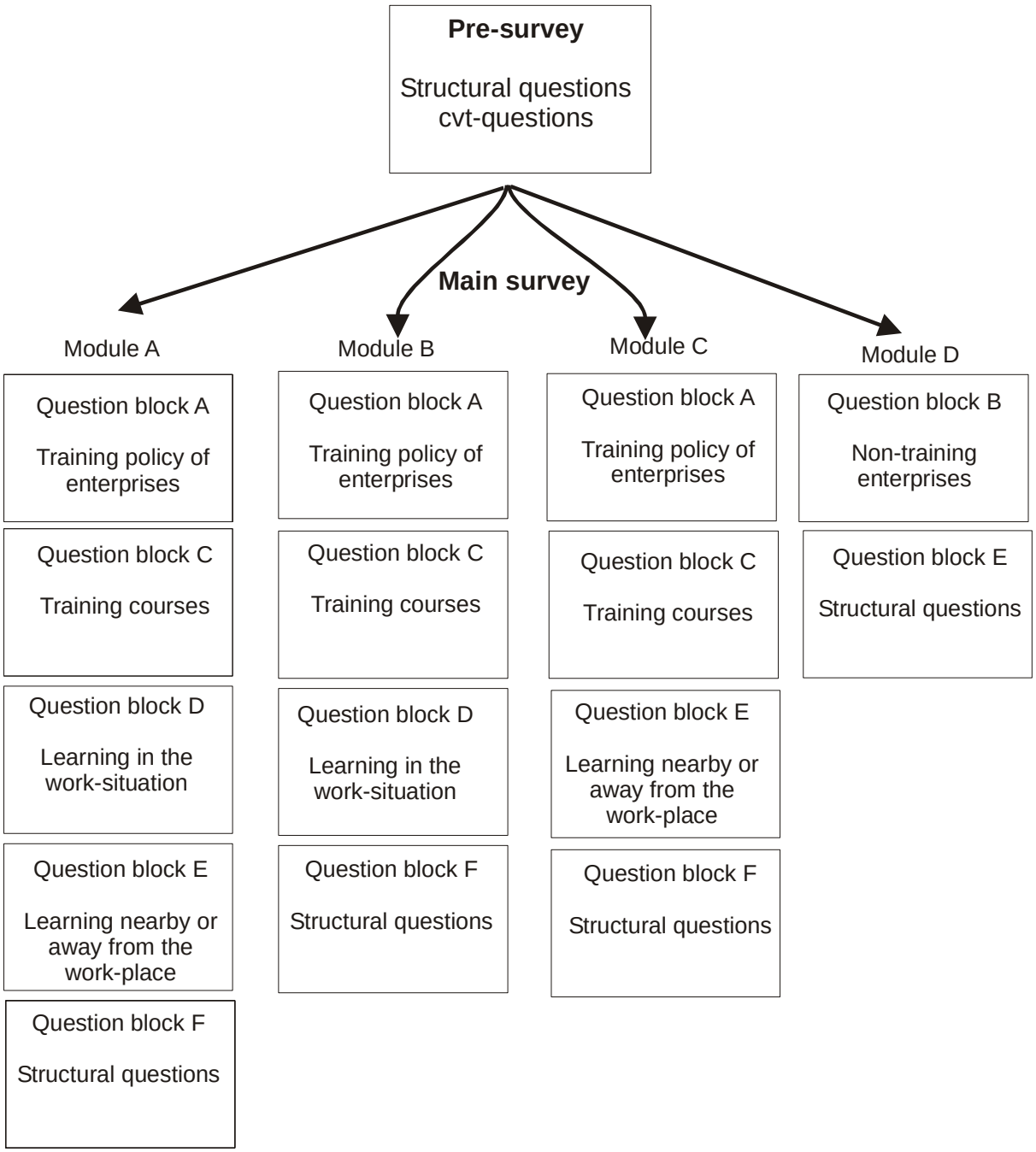
Role of interviewer

The information collected in the first stage will be integrated into the main questionnaire, independent of the form of the data collection. This will give the contact person the impression that the interviewer (using e.g. CAPI/CATI) already is informed about the enterprise. However, the training of interviewers is of vital importance for the data collection in the main survey phase.

5. Example of a staging of the survey and modularisation of the questionnaire approach

The following graph is an example of how to implement the staging/modularisation approach for CVTS. Here we exemplify the staging/modularisation approach used by Germany for CVTS1.

Figure 1: Staging of the survey and modularisation of the questionnaire of CVTS1 in Germany (1994/1995)



The pre-survey of the German CVTS1 consisted of eight questions in total. In addition to three structural questions (enterprise size, economic sector, number of local units) and a question about data access in the enterprise, there were four questions about continuing vocational training (provision of training courses, division of training courses into internal/external courses, a general question asking participation in training courses and provision of "other forms" of training). The modular main survey (divided in 4 modules) included the following question blocks: training policy of enterprises, questions to non-training enterprises, training courses, "other forms" of continuing vocational training - Learning in the work-situation, "other forms" of continuing vocational training - learning nearby or away from the workplace and structural questions.

6. Summary

Staging has been carried out by a number of countries in CVTS1 and or CVTS2. Some countries have been satisfied with the process and believe that it helped them achieve better results. The example of Statistics Austria is very positive in this regard. Although Statistics Austria implemented the survey much later than the other CVTS2-countries, the complete survey, starting with the fieldwork up to the final data set, was accomplished in eight months. The main advantage of this approach according to the paper of Statistics Austria is the efficiency and flexibility of the approach, better collection of data of non-training enterprises, a better identification of the unit non-response and minimizing the response burden for enterprises.

However, other countries are less positive and some have not used staging in CVTS. Statistics Sweden have presented the main arguments against staging along with a discussion of possible advantages. The paper of Statistics Sweden is more sceptical of the use of a staging/modularisation approach. They especially stipulate the methodological disadvantages of this approach. It is commonly assumed that staging might enhance response rates, but they argue that it depends on different factors if this advantage will be put into effect: the mandatory or voluntary nature of the survey, the way of data collection, obligatory ex ante information to the enterprises and responding behaviour of enterprises (culturally divergent) are amongst the determining factors.

In view of these different arguments, and the differing circumstances that apply in each country, the CVTS3 M Project Team does not believe that it is appropriate to make a firm recommendation in favour or against staging. Rather, it recommends that each country examines the issue from its own circumstances. In conducting this examination, the countries should consider the factors outlined above including likely response rates, the percentage of non-training enterprises, their expected data collection methods, and availability of resources to minimise non-response. Countries should make a clear decision on which method they intend to use and ensure that the implications of their decision for sampling, cost, methodology and weighting are fully considered. Their proposed methods should form part of their proposal to, and agreement with, Eurostat and the European Commission.

Annex 1:

Experiences with Staging and Modularisation in conducting CVTS2 in Austria

Expertise by

Statistics Austria

Introduction

This report outlines the socio-historical context and its impact on survey design, which eventually led to the use of staging and modularisation as central elements of the survey. We shall also attempt to review the efficacy of selected measures by drawing on administrative data from the data collection phase. Finally, we shall discuss the effects of staging on the calculation of errors.

Original situation

Due to the fact that the survey on continuing vocational training (CVT) in Austria was commissioned later than originally planned, practical preparations were delayed from the beginning of 2000 to December of the same year. As the completion/delivery date (envisaged for end of 1st quarter 2001) remained unchanged, the survey had to be conducted within a significantly shorter period.
This situation prompted us to develop a design that deviated from the original specifications. The design was optimised with a view to the following key facts:

- the survey would be voluntary
- the survey would take optimal account of respondents' needs
- the survey would be conducted in a timely manner with no time cushions provided.

The first two aspects are related to the legal context applying to official statistics in Austria; these are of central importance to the survey's acceptance, which in turn is a key factor determining timely implementation.
To improve acceptance, surveyed enterprises were to be assured of the following:

The survey would abstain from re-collecting structural data for economic statistics which are already collected during mandatory surveys.

The questionnaire would be reduced to an absolute minimum, both as far as the number of questions and the overall length of the tool are concerned.

In addition, the first assurance was also based on the consideration that larger enterprises would appoint different personnel to respond to the questions on CVT activities and to those on hours of work or total labour cost, which again implies extra time effort for these businesses.
Generally speaking, the fact that questionnaires might "go astray" in the companies was deemed a potential risk which would have to be addressed by immediate follow-up.
A working group was established to exchange experience concerning voluntary surveys - a concept little used for economic statistics in Austria. The key element in this strategy is to contact respondents by telephone. The calls would announce a written poll that is due to follow, and motivate people to respond. It must be strictly distinguished from reminder calls, queries or interviews over the telephone etc. Though entirely based on know-how from personal surveys, this know-how was largely used as a basis for designing CVTS2.
In fact, drawing from the stock of addresses included in the business register and the contact data collected from statistical business surveys (short-term statistics, structural business statistics), the procedure should be as follows: a **first stage** (preliminary survey, screening) should identify the persons to be contacted for CVT activities and the information on basic activities undertaken in the year under review which is required to plan and gear the main survey. This would be effected by a written survey supported by extensive telephone support. It would only be in the **second phase** (main survey) that the detailed questionnaire, which is geared to the specific activities identified in each case, would be transmitted to the relevant

persons in the enterprises concerned. This stage would also make ample use of motivation calls encouraging enterprises to take part in the written survey.
The above are some basic considerations which have led us to apply staging and modularisation in designing CVTS2.

Impact of staging

Feedback was one of the critical aspects anticipated from the outset: We doubted whether a response rate of 50% would be at all realistic for a voluntary survey that includes a relatively comprehensive set of questions. On the other hand, it was obvious that for all intents and purposes, an even much higher participation by enterprises would be required to obtain valid results. Apart from potential distortions caused by systematic drop-outs, massive losses of answers are generally suited – especially in strata with low case numbers in the first place (e.g. the very large enterprises) – to raise questions concerning the quality of estimates.
At any rate, it was due to the staging of the survey that essential information already collected during the screening stage should receive broad feed-back. By way of example, the percentage of CVT, i.e. the proportion of enterprises engaged in continuing vocational training activities in the year under review, has partly been computed on the basis of data provided by more than 90% of the sampled enterprises (see table 1).
Apart from applying staging, another measure to assure quality was to define an intensive follow-up sample which makes a point of providing extensive telephone support during the data collection stage. This sub-sample covered all enterprises employing from 250 staff, while the sampling fraction was reduced as size class decreased. This ensured extensive follow-up of large enterprises, which are of key importance particularly for quantitative dimensions, while small-sized enterprises were relieved under the intended policy of ”sparing” respondents.
At the same time, the sub-sample allowed concentration of limited call centre resources. When comparing the returns of the preliminary screening, it is obvious that the "10 to 19 employees" category is under-represented in the total sample (53 % response rate as compared to 85% in the intensive follow-up sample). The difference is primarily due to time and financial constraints, causing only 31% of this size class to be contacted by telephone, while almost two thirds of the intensive follow-up sample were followed up over the telephone. Accordingly, the reduced willingness of smaller enterprises to respond at this stage is much less dramatic (85% compared to 90% for companies employing 20 to 49 staff) than it may seem at first glance.
Transmitted in the form of a conventional written survey, the short questionnaire was answered by approx. 30% of the enterprises. Of those not approached by telephone, the questionnaire was returned by a quarter of the enterprises employing from 10 to 19 staff, by approx. 30% of those employing between 20 and 249 staff, and by slightly more than a third of the enterprises employing from 250 staff. Given that the subsequent procedure was described in a letter accompanying the short questionnaire, no further follow-up was made at this point. On the other hand, the enterprises contacted by telephone had the opportunity to immediately answer the short questionnaire over the telephone, obviating the need for written processing and return by surface mail.
Upon completion of the screening exercise, 78% of the sampled enterprises, or 90% of those taking part in the intensive follow-up sample, had provided basic information on CVT activities. Concerning the relevant contact persons, nearly two thirds of enterprises reported updates - 23% had appointed another person, for whatever reason (fluctuation, shift in responsibilities etc.), while 39% mentioned a name for the first time. A quarter of the contacts remained unchanged; in 13% of the cases, no contact was entered even after the screening exercise (see table 2).

As expected, there is a correlation with size. Small enterprises tend to be more stable (no entry or same contact person), while larger enterprises more often reported change (different or new contact). Contact data were changed for more than 80% of the enterprises with staff from 250, compared to as much as nearly one half of the enterprises employing between 10 and 19 staff.
If we exclude the enterprises not indicating any CVT activities in the year under review, contact data were recorded for the first time in 58% of the enterprises employing from 10 to 19 staff, whereas a total of almost 70% of the data received from this size class were updated. In this exercise, the number of enterprises indicating no contact person was reduced to one half, totalling only 6%.
The positive development of the screening stage seems to be due, on the one hand, to the **conciseness of the screening questionnaire**, which indeed included only the following questions: participation in internal or external vocational training courses and other forms of continuing vocational training (Yes/No answers); any reasons for absence of CVT activities, if applicable (8 categories of answers); and contact data relating to a relevant person in the company. Therefore, responding to the short questionnaire involved a minimum of effort.
On the other hand, we were able to rely on a call centre – at that point an ad-hoc facility staffed by in-house volunteers – supported by a highly motivated team of initially ten. By assigning a defined set of addresses to a specific set of colleagues, our telephone contacts were generally characterised by consistency and continuity. The only discontinuity occurred as some in-house staff fluctuation arose during the course of the survey.
The records at our disposal allow no systematic conclusions as to the quality of telephone contacts (i.e. availability of contact, accessibility through secretariat only etc.) nor do they provide any details on frequency. Generally the conversations with respondents were described as matter-of-fact and occurring in a positive climate – both during the preliminary screening phase and when motivating follow-up calls were made. However, in many cases the relevant persons were contacted a number of times, particularly during the main survey.
An actual matter of concern is the disappointing rate of responses received during Stage 2, when enterprises were faced with the challenge of completing complex questionnaires. Though no problem occurred in placing the previously announced questionnaires in the companies, only a total of 40% of the enterprises engaged in CVT activities actually responded to the main survey questionnaire. In this context, we observed no trend in the response rates as far as size class is concerned (see table 3).
Overall, the pattern of telephone contacts suggests a rather modest level of success of motivation calls placed in the context of a survey of enterprises. The response rates achieved in the three size classes of up to 249 employees were only minimally higher in the intensive follow-up than in the total sample, whereas they differed only slightly from smaller enterprises in the 250+ size classes, which were followed up with particular intensity. However, this does not imply that telephone support was indeed redundant during data collection – in which case we can only speculate as to the response rates achieved for large enterprises. On the contrary, we even believe – and in fact, rely on some confirmed cases – that personal contact frequently fostered engagement and commitment from relevant persons in large enterprises – which indeed prevented the survey's response rate from declining. On the other hand, it seems that a similar result has been achieved for smaller enterprises not covered under the intensive follow-up sample, which received conventional reminders by surface mail.
However, one of the striking findings is that although nearly all enterprises covered by the intensive follow-up sample (but which eventually dropped out in the main survey) were contacted, it was not possible to improve on the response rate: 90% to 95% of no-response enterprises had effectively been followed up by motivation calls.

Modularisation

It was envisaged to use tailor-made questionnaires to accommodate the companies to the extent possible and to enhance efficiency so as to obtain higher response rates.
The modularisation criteria were as follows:

- CVT activities according to the screening survey
  Depending on the activity, enterprises were supposed to receive the block of questions concerning CVT courses only (Section C), other forms of CVT (Section D) or both.
- Participation in economic statistical surveys
  Depending on enterprises' classification according to NACE, economic statistical data were available from mandatory surveys which could be merged to survey data via an identifier on the Business Register.
- Company size
  For enterprises with 10 to 19 employees, general questions from sections B and C (B1 to B7, C7 and C8) were to be eliminated as their content would often not apply directly to small enterprises. Another intention was to preclude problems of acceptance (i.e. the point often raised as to statistical surveys' excessive burden for small enterprises).

These three dimensions were combined to prepare 12 different questionnaires - one to two A3-size sheets each - specifically tailored to the respective target groups. The questionnaires were also delivered in electronic format for online completion (transmission by e-mail as per enterprises' request made during the preliminary survey).
In line with the screening survey, 38% of valid screening cases received a questionnaire on continuing vocational training taking the form of courses (types 1 to 4), while 42% received questionnaires concerning courses and other forms of continuing vocational training (types 5 to 8); 1% received a questionnaire about "other forms of continuing vocational training" only. Almost one fifth of the enterprises were spared the main survey as the reasons for "no continuing vocational training" were already identified in the screening exercise and no further questions were required from Section B.
The following questionnaire types were used (as a percentage of all questionnaires, i.e. excluding those enterprises which had not stated any CVT activity):

16% received the short version for small-sized enterprises (types 1 and 5), which did in fact contain questions concerning hours worked and total labour cost.

21% received questions concerning continuing vocational training courses (Section C), excluding "other forms of CVT" as well as questions on structural economic data (A3 and A4) (type 2).

14% were neither asked to submit responses on Section D nor on the question concerning total labour cost (A4) (type 3).

At 23%, type 6 was most frequently used; similar to type 7, this questionnaire covered courses and other types of CVT, while questions A3 and A4 were eliminated.

Type 7: 22%, including questions on hours worked.

Questionnaire type 12 was not used at all, while the other three types were only used sporadically and only asked about "other forms".
In some cases, enterprises stated having made a mistake in the screening survey and not being able to report any CVT activity for 1999. In the opposite case – e.g. companies that indeed should have reported other types of CVT – there was no change due to the design.
There are only minor differences in response rates according to type of questionnaire (see table 4):

Types 1 and 2, which abstained from certain questions included in type 3, scored a 7 percentage points increase in the response rate over type 3; however, this gap becomes smaller in the intensive follow-up sample.

This situation is similar for types 5 and 6 versus type 7, though with a smaller gap in the first place (4 percentage points).

The questionnaires only asking about CVT courses (types 1 to 3) show slightly higher response rates than the comparable questionnaires asking about courses and "other forms" (types 5 to 7), though the gap amounts to a maximum of only 5 percentage points.

Though based on low case numbers as only 10% of randomly selected enterprises with 10 to 19 employees received a long version of the questionnaire (types 2, 3, 6 or 7), it is nevertheless surprising that small enterprises' acceptance of types 1 and 5, i.e. those variants that took particular account of respondents' needs, was indeed not superior to the "full-fledged versions" (cf. table 5).

Analysed by size class, we have also found that the questionnaires asking about hours worked (types 3 and 7) generally resulted in lower response rates than those (types 2 and 6) which only differed by eliminating this question.

In conclusion, these observations lead us to believe that cutting down on easily answerable, factual questions had barely any positive impact on willingness to respond; however, that even minor changes in a complex question triggered significant effects. Indeed, this finding corroborates our overall impression of the survey in that the intensive follow-up sample was largely exploited by motivation calls and that the feedback could not be expected to improve significantly even by extending deadlines. It seems that the critical point correlated with the effort required to answer the quantifying questions, which in the majority of cases obviously outweighed the expected benefit by far.

Given the time constraint in conducting the survey, we never considered summarising the results of the preliminary screening survey in a short brochure nor to disseminate it to the businesses concerned as a means of promoting the main survey.

We also note that the preliminary screening results had a considerable impact on the final result. Published results on a wider European scale suggest that by strictly linking the other forms of continuing vocational training to a simplistic screening question, the overall prevalence of these CVT measures was obviously grossly underestimated. Instead of surveying the five different forms on an individual basis, the screening sheet only listed measures in combination with a request for providing a summary answer. By strictly dividing the question groups on CVT courses and "other forms" in different modules, it proved impossible to make adjustments at a later stage based on feedback received during the survey phase.

Calculation of sampling errors

For this survey, as regards the estimators under consideration, we relied on standard formulas for stratified random sampling. For the quality report in particular, we had to calculate coefficients of variation for the following estimators:

1. Sum of employees
2. Sum of employees in enterprises providing continuing vocational training
3. Sum of enterprises providing CVT
4. Share of enterprises providing CVT
5. Sum of employees having taken part in CVT courses
6. Proportion of employees having attended CVT courses in total sum of employees
7. Proportion of employees having attended CVT courses in total sum of employees in enterprises providing CVT
8. Sum of expenditure on CVT courses

All these estimators, hence their coefficients of variation, had to be calculated for employee size class in combination with pre-defined NACE 2-digit groups.
It follows from the estimator characteristics that the following standard formulas could be used for the coefficient of variation:
To estimate proportions (4.)

$$v(\hat{A}) = \frac{N(N-n)}{n-1} p(1-p) \ ,$$

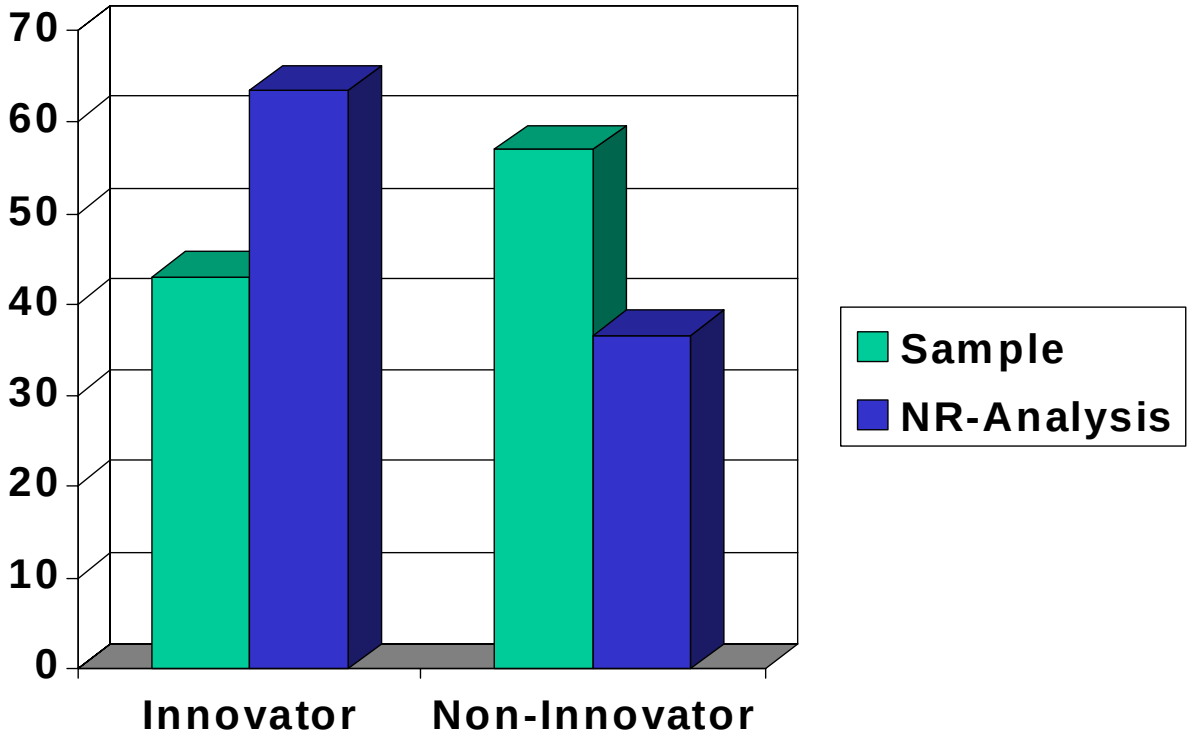
for ratio estimators (6. - 7.) $\hat{R} = \frac{\ddot{X}}{\ddot{Y}}$

$$v(\hat{R}) = \frac{N-n}{n^2 \ddot{X}^2} \frac{\sum_{i=1}^{n} (y_i - Rx_i)^2}{N-1}$$

and for totals (1.-3., 5.,8.)

$$v(\hat{Y}) = \frac{N^2 s^2}{n} \frac{N-n}{N} \ ,$$

with s^2 representing the variance estimated from the sample.
The formulas were applied to all strata. This proved sufficient as the stratification corresponded exactly with the analytic variables. For the calculation of errors of the marginal distribution the values for the respective strata were accumulated.
The two-stage survey procedure was considered in the calculations of variance only to the extent that different sample sizes (small n) were taken into account for individual estimators. Staging basically only prevented some of the enterprises considered active in CVT from refusing collaboration because of the length of the remaining questionnaire (in case of a positive filter). Therefore we cannot refer to a two-stage sampling scheme in the strict sense, but rather to a homogeneous stratified random sample for calculation of variance (estimates).
Based on Statistics Austria experience with a non-response analysis in connection with the 2000 Community Innovation Survey (CIS3) - a voluntary survey of approx. 3,700 units of enterprises - the following example illustrates the effect triggered by a questionnaire design starting with a filter question. The introductory question was to the tune of “Is your company an innovator?”. Overall, unit non-response amounted to approx. 50%. When approx. 10% of the deniers were subsequently analysed in a non-response analysis – they were only asked the question concerning the innovation activity (i.e. the reverse procedure to CVTS2) – the estimation of innovation companies resulted as follows:



This graph clearly shows that the sampling results underestimate the number of innovators. The results of the non-response analysis required corresponding modification of the weighting for CIS3, which increased the variances of estimators.
Not conducting CVTS2 in two stages would probably have resulted in distortions as well, or, after a similar NR analysis, in the need to increase variation of the weighting factors.

Summary

Staging and modularisation were used as a strategy to address a specific given situation and were applied to a written survey combined with massive telephone support provided during the course of the survey.
Thanks to telephone contacting, the first stage of the survey recorded a 90% or even higher rate of return in those strata immediately covered by follow-up whenever enterprises failed to respond.
Though the names of relevant persons were updated in the first stage and emphasis on telephone coverage continued through the second stage of the survey, we did not succeed in motivating respondents to take part in the main survey to a similar degree as it was possible during the preliminary screening exercise.
Although the specific questionnaires were tailored to the businesses concerned and emphasis was placed on accommodating respondents to the extent possible, the response rate achieved for the main survey only amounted to 40%. Nevertheless, it can safely be assumed that staging and telephone support significantly contributed to the overall success in achieving response.
The results of the screening exercise were used to fine-tune the main survey, in addition to modifying the weighting scheme so as to avoid non-response-related distortions in the distribution of CVT and non-CVT enterprises.
Not conducting the survey in stages would probably have raised the need for a non-response analysis, ultimately also resulting in considerable modification of the weighting scheme.

Tables

Table 1a: Response in the screening stage - total sample

Size class (Employees) *)	Gross sample	of which		Screening response rate	of which	Percentage of enterprises contacted by telephone
		contacted enterprises	of which valid after screening		Response rate without telephone contact	
	absolute			in %		
less than 10	87	87	87	100,0	48,3	51,7
10 to 19	2.230	2.198	1.173	53,4	25,8	30,5
20 to 49	1.404	1.394	1.208	86,7	30,0	61,8
50 to 249	2.268	2.253	2.011	89,3	29,2	64,3
250 to 499	512	509	463	91,0	33,6	61,5
500 to 999	238	236	219	92,8	38,1	58,9
from 1000	169	166	159	95,8	35,5	60,8
Total	6.908	6.843	5.320	77,7	29,3	52,3

*) Allocation according to empirical number of employees in responding enterprises.

Table 1b: Response in the screening stage - intensive follow-up sample

Size class (Employees) *)	Gross sample	of which		Screening response rate	of which	Percentage of enterprises contacted by telephone
		contacted enterprises	of which valid after screening		Response rate without telephone contact	
	absolute			in %		
less than 10	46	46	46	100,0	28,3	71,7
10 to 19	750	741	632	85,3	24,7	65,9
20 to 49	698	694	622	89,6	28,2	65,9
50 to 249	1.150	1.147	1.036	90,3	29,9	64,9
250 to 499	498	495	449	90,7	33,5	61,4
500 to 999	235	233	216	92,7	37,8	59,2
from 1000	168	165	158	95,8	35,8	60,6
Total	3.545	3.521	3.159	89,7	29,8	64,3

*) Allocation according to empirical number of employees in responding enterprises.

Table 2a: Relevant contact persons in enterprises after the screening exercise

Size class (Employees) *)	Contact person in enterprise				Valid screening cases
	none	same	other	new	
	row percent				absolute
less than 10	33,3	16,1	3,4	47,1	87
10 to 19	29,0	23,5	8,0	39,5	1.173
20 to 49	15,4	32,5	16,9	35,3	1.208
50 to 249	6,0	25,8	29,4	38,8	2.011
250 to 499	2,2	12,5	44,9	40,4	463
500 to 999	1,4	17,8	31,1	49,8	219
from 1000	1,9	8,8	34,0	55,3	159
Total	13,0	24,6	23,0	39,4	5.320

*) Allocation according to empirical number of employees in responding enterprises.

Table 2b: Relevant contact persons in training enterprises after the screening exercise

Size class (Employees) *)	Contact person in enterprise				Valid screening cases
	none	same	other	new	
	row percent				absolute
less than 10	33,3	16,1	3,4	47,1	87
10 to 19	12,8	17,8	11,5	57,9	729
20 to 49	7,2	27,8	21,5	43,5	902
50 to 249	2,9	24,0	31,3	41,7	1.831
250 to 499	2,0	11,9	45,6	40,5	452
500 to 999	0,9	17,7	31,6	49,8	215
from 1000	1,9	8,8	34,0	55,3	159
Total	5,8	21,5	27,0	45,6	4.375

*) Allocation according to empirical number of employees in responding enterprises.

Table 3a:
Response in the main survey (excluding enterprises not providing CVT)
Total sample

Size class (Employees) *)	Question-naires despatched	of which	Response rate in main survey	Percentage of enterprises receiving motivation calls	of which	
		with valid answer			Percentage among responding enterprises	Percentage among missed answers
	absolute		in %			
less than 10	87	51	58,6	13,8	19,6	5,6
10 to 19	729	295	40,5	39,0	24,4	48,8
20 to 49	902	386	42,8	44,7	31,1	54,8
50 to 249	1.831	682	37,2	49,6	33,9	59,0
250 to 499	452	175	38,7	82,1	59,4	96,4
500 to 999	215	100	46,5	77,7	63,0	90,4
from 1000	159	63	39,6	85,5	71,4	94,8
Total	4.375	1.752	40,0	52,2	36,8	62,4

*) Allocation according to empirical number of employees in responding enterprises.

Table 3b:
Response in the main survey (excluding enterprises not providing CVT)
Intensive follow-up sample

Size class (Employees) *)	Question-naires despatched	of which	Response rate in main survey	Percentage of enterprises receiving motivation calls	of which	
		with valid answer			Percentage among responding enterprises	Percentage among missed answers
	absolute		in %			
less than 10	46	27	58,7	26,1	37,0	10,5
10 to 19	390	156	40,0	71,0	42,9	89,7
20 to 49	476	220	46,2	71,0	43,6	94,5
50 to 249	946	375	39,6	75,1	46,7	93,7
250 to 499	438	161	36,8	84,0	62,7	96,4
500 to 999	212	97	45,8	78,8	64,9	90,4
from 1000	158	62	39,2	86,1	72,6	94,8
Total	2.666	1.098	41,2	75,3	50,7	92,5

*) Allocation according to empirical number of employees in responding enterprises.

Table 4a:
Response in main survey (excluding enterprises not providing CVT) according to
Type of questionnaire - total sample

Type of questionnaire	Question-naires despatched	of which	Response rate in main survey	Percentage of enterprises receiving motivation calls	of which	
		valid answer			Percentage among responding enterprises	Percentage among missed answers
	absolute		in %			
Type 1	456	203	44,5	36,0	20,2	48,6
Type 2	931	415	44,6	49,5	31,1	64,3
Type 3	616	230	37,3	51,8	32,6	63,2
Type 4	20	9	45,0	60,0	55,6	63,6
Type 5	240	95	39,6	37,9	25,3	46,2
Type 6	1.010	413	40,9	62,4	48,4	72,0
Type 7	952	344	36,1	58,0	43,9	66,0
Type 8	39	13	33,3	64,1	46,2	73,1
Type 9	11	4	36,4	72,7	100,0	57,1
Type 10	25	18	72,0	48,0	38,9	71,4
Type 11	16	8	50,0	50,0	37,5	62,5
not known	59	-	0,0	.	.	.
Total	4.375	1.752	40,0	52,2	36,8	62,4

Table 4b:
Response in main survey (excluding enterprises not providing CVT) according to
Type of questionnaire - intensive follow-up sample

Type of questionnaire	Question-naires despatched	of which	Response rate in main survey	Percentage of enterprises receiving motivation calls	of which	
		valid answer			Percentage among responding enterprises	Percentage among missed answers
	absolute		in %			
Type 1	230	98	42,6	70,9	40,8	93,2
Type 2	577	256	44,4	72,4	43,8	95,3
Type 3	387	155	40,1	73,6	42,6	94,4
Type 4	13	7	53,8	84,6	71,4	100,0
Type 5	129	53	41,1	68,2	43,4	85,5
Type 6	670	279	41,6	79,3	60,6	92,6
Type 7	586	226	38,6	79,7	55,8	94,7
Type 8	28	9	32,1	85,7	66,7	94,7
Type 9	8	4	50,0	100,0	100,0	100,0
Type 10	12	9	75,0	66,7	55,6	100,0
Type 11	6	2	33,3	83,3	50,0	100,0
not known	20	-	0,0	.	.	.
Total	2.666	1.098	41,2	75,3	50,7	92,5

Table 5a:
Questionnaires despatched for the main survey, by type and size of enterprise
Total sample

Type of Questionnaire	Size class (Employees) *)							
	less than 10	10 to 19	20 to 49	50 to 249	250 to 499	500 to 999	from 1000	Total
	absolute							
Type 1	38	402	14	2	-	-	-	456
Type 2	2	34	314	443	94	34	10	931
Type 3	-	26	169	343	47	20	11	616
Type 4	-	-	3	10	5	2	-	20
Type 5	12	216	9	1	2	-	-	240
Type 6	1	13	157	512	174	85	68	1.010
Type 7	-	15	198	482	120	70	67	952
Type 8	-	2	5	16	10	3	3	39
Type 9	-	11	-	-	-	-	-	11
Type 10	-	2	15	8	-	-	-	25
Type 11	-	-	4	12	-	-	-	16
not known	34	8	14	2	-	1	-	59
Total	87	729	902	1.831	452	215	159	4.375

*) Allocation according to empirical number of employees in responding enterprises.

Table 5b:
Response rate for main survey, by type of questionnaire and size of enterprise
Total sample

Type of Questionnaire	Size class (Employees *)							
	less than 10	10 to 19	20 to 49	50 to 249	250 to 499	500 to 999	from 1000	Total
	in %							
Type 1	100,0	37,1	100,0	100,0				44,5
Type 2	100,0	79,4	46,2	42,9	37,2	41,2	20,0	44,6
Type 3		69,2	39,6	34,1	27,7	65,0	18,2	37,3
Type 4			66,7	20,0	80,0	50,0		45,0
Type 5	91,7	33,3	100,0	100,0	100,0			39,6
Type 6	0,0	92,3	37,6	38,3	43,1	49,4	42,6	40,9
Type 7		66,7	38,4	32,8	35,0	41,4	43,3	36,1
Type 8		50,0	20,0	31,3	40,0	33,3	33,3	33,3
Type 9		36,4						36,4
Type 10		100,0	66,7	75,0				72,0
Type 11			75,0	41,7				50,0
not known	0,0	0,0	0,0	0,0		0,0		0,0
Total	58,6	40,5	42,8	37,2	38,7	46,5	39,6	40,0

*) Allocation according to empirical number of employees in responding enterprises.

STATISTICS SWEDEN
Department of labourmarket and
education statistics

Annex 2:

Pros and cons concerning staging in CVTS3

by

Statistics Sweden

General comments

One aim of using staging in the CVTS is in a first stage to present the sampled enterprises with a short questionnaire where questions are asked mainly in order to get some basic and vital information about the enterprise. The received information should make it possible to decide later on which tailor made questionnaire should be sent to the enterprise in the second stage. The separation into different questionnaires in the second stage could for instance be whether an enterprise has internal courses, external courses, IVT, CVT or other forms of training. As the questionnaire in the first stage is intended to contain just a few questions it is possible at the same time for instance to ask necessary questions to non-training enterprises.

One should be aware of the fact that in statistical theory the staging approach proposed in CVTS actually means a two-phase sampling design. This means that other statistical theories will be applicable than in a more common one-phase survey. Furthermore, weighting procedures have to be adapted.

In some way one could say that one basic intention with a staging approach is to avoid scaring off the respondents with a questionnaire with a lot of questions. In other words when using Internet based data collection, telephone interviews or face-to-face interviews with CATI or CAPI this reason does not exist, as the respondent with those kinds of data collection methods only will get the questions relevant for them anyhow.

When using postal questionnaires as the data collection method, staging may be an option subject to the condition that one is quite certain of getting a high response rate in the first stage and an overall quite good total response rate for both stages.

For countries with a quite high percentage of enterprises providing different kinds of training staging involves considerably less advantages or even disadvantages than for countries where the use of one kind of training usually means much less occurrence of other kinds of training. The possible advantages of staging are also higher for countries with a lot of non-training enterprises as questions to non-trainers should be included in the first stage.

A staging approach may perhaps be something that one considers if one is afraid of getting too low total response rate with a one-stage approach.

Pros

1. By using a short questionnaire you can often expect to get quite high response rates in the first stage.

2. It will be possible to present early results for the questions included in the first stage. However, in order to be able to produce interesting preliminary results from the first stage, you will probably have to include a few rather burdensome questions like for instance the number of course participants already in the first stage.

3. If it is possible to get high response rates in the first stage it may be possible to use that information in the weighting procedure in the second stage. However, to be able to do that some questions must be included in the first stage otherwise we will not be able to improve the quality of the estimates in the second stage. These questions are rather burdensome for the enterprises like for instance the number of course participants.

4. A staging approach may be a clear alternative for countries with a quite small proportion of training enterprises. It will probably be possible to get higher response rates and better information for non-trainers if we include questions to this group of enterprises in the first stage than we would get in a one-stage survey. For countries with a very high proportion of non-trainers a staging approach may even lead to a somewhat higher total response rate. However, in those cases one should also be prepared for the risk of getting a somewhat lower response rate for trainers.

Cons

1. The procedure with a staging approach means that non-response units in the first stage will not get the questions intended for the second stage. It is therefore of vital importance to get a very high response rate in the first stage. In our opinion the expected response rate in the first stage should be at least 70 per cent before considering a staging approach. It is therefore very important to allocate resources to reminders and also to put aside a rather long period of time for the data collection especially in the first stage.

2. It is important to get a good estimate of the probability to answer in the first stage. In order to be able to do that one has to base the estimate on some model assumptions. It will be much more difficult to make necessary assumptions if the response rate in the first stage is low. All countries will probably have access to information about the 20 NACE and the three size classes used in CVTS2 for both the sample allocation and for the presentation of results. However, if the response rate is rather low in the first stage, one will probably also need information for more detailed size classes and probably also for other variables in order to have a model good enough for estimating the probability to get an answer in the first stage. One simply has to know what characterises non-response units in the first stage.

3. According to statistical theory the variance in results will be bigger in a two-phase sampling design than in a one-phase sampling design. It will not be possible to improve the quality in the estimates based on the second stage, if we do not have access to information that can be used as auxiliary information in the first stage.

4. A staging approach will be quite dependent on receiving correct answers in the first stage. An enterprise that for different reasons give incorrect answers to some of the questions in the first stage will not get the right questionnaire in the second stage or will not at all get any questions in the second stage. In a one-stage survey it is quite often possible to see whether a respondent may unintentionally have ticked the wrong answer to a certain question. That will not be possible to do in a staging approach. Incorrect answers in the first stage are something that we will have to cope with, as we normally cannot do anything about it.

5. A staging approach is mainly only an option to consider under certain circumstances when using a postal questionnaire.

6. In some countries like for instance in Sweden there exists special authorities safeguarding the response burden on enterprises. To approve of the survey these authorities can demand that the enterprises must be informed that they may get another questionnaire later on and also that one has to describe in short what kind of

questions that will be included in the second questionnaire. In such cases a clear risk exists that some enterprises may decide already in the first stage not to participate at all in the survey.

Conclusion

Advantages of staging are more likely

- When using postal questionnaires
- Under the condition of high response rates in the first stage (at least 70 %)
- Provided that time and money allow for intensive re-contacting to avoid unit non-response in the first stage
- In countries with a high proportion of non-training enterprises, if questions concerning non-trainers are included in the first stage
- In countries with training enterprises focussing on one kind of CVT and usually not providing any other kind of training

Staging is connected with some disadvantages and risks:

- Staging reduces the possibilities of correcting answers received in the first stage
- A staging approach implies a two-phase sampling design. This requires application of other statistical theories than in a more common one-phase survey and adaptation of weighting procedures. Variance in results will be higher than in a one-phase sampling design, too.
- Non-response and incorrect answers in the first stage are impacting on the second stage. E.g. delivering the correct version of the questionnaire in stage 2 is contingent on obtaining correct answers in stage 1.

It is assumed that staging might enhance response rates, but it depends on different factors if this advantage will be put into effect: The mandatory or voluntary nature of the survey, the way of data collection, obligatory ex ante information to the enterprises and responding behaviour of enterprises (culturally divergent) are amongst the determining factors. Taking into consideration that advantages are contingent on method of data collection, time and money allocated at re-contacting in the first stage, and distribution of training activities across enterprises, it might be favourable not to regulate staging. Instead, it might be advantageous to treat staging as a matter of countries' discretion.

STATISTICS SWEDEN
Department of labourmarket and
education statistics

Annex 3:

Two-phase sampling approach in CVTS3

by

Statistics Sweden

Introduction

It has been discussed to use a staging approach in the CVTS3. This means that the enterprises in the sample will receive a short questionnaire with basic questions in the first stage of the survey. Depending on the responses to the questions in the first stage, the enterprises will receive different questionnaires in the second stage of the survey. One possibility would be to distinguish between trainer enterprises and non-trainer enterprises in the first stage of the survey.

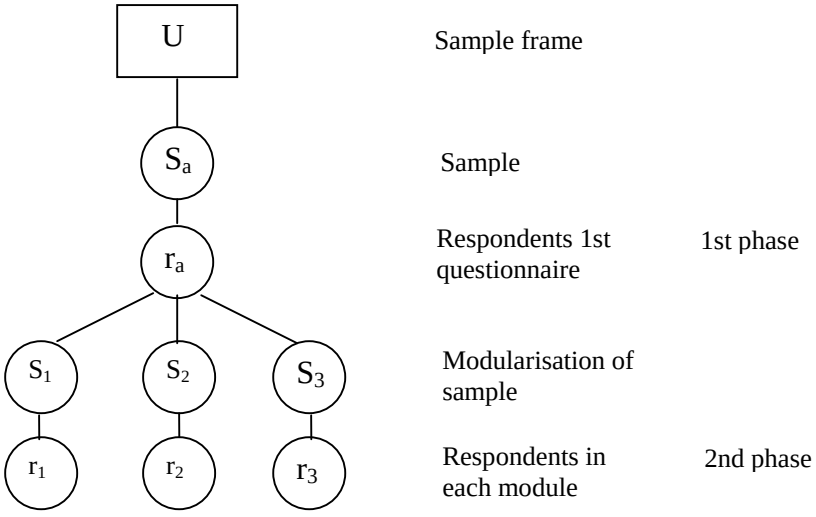
Staging of the survey means in practice that the statistical theory for two-phase sampling should be applied when calculating the estimates for the confidence intervals. This is actually the case in the presence of non-response for a single-phase sample as well. In such cases the variance should reflect both the uncertainty caused by the selection of a sample and the uncertainty caused by the non-response, i.e. a variance component for the selection of a sample and a variance component for the non-response.

Non-response will occur for the questionnaire in the first stage of the survey and for the different questionnaires in the second stage as well. This means that it is even more important to use the theory for two-phase sampling when staging of the survey is used instead of a single stage survey. The purpose with this paper is to describe how two-phase sampling can be applied when staging of the survey is used.

Staging of the survey is normally called split questionnaires in the statistical literature. Professor Jan Wretman has for example written a research report about estimation in sample surveys with split questionnaires.

Selection of the sample

The figure below tries to give a general picture of the connection between staging of the survey and two-phase sampling.



Actually this can be seen as a three phase sample where S_a is the first phase sample and the respondents on the 1st questionnaire, r_a, can be seen as the second phase sample and the respondents in each module (r_1, r_2 and r_3) can be seen as the third phase sample. This can however be approximated as a two-phase sample, which is shown in the figure above.

Even if no sample has been drawn and questionnaires are addressed to the total population, the respondents are actually selected from the sample by some selection mechanism. The probability to respond is however unknown. It is often assumed that the non-response within strata is randomly distributed. The weight for each responding enterprise will depend on the sample design and on the non-response.

Estimation of confidence intervals

The length of the confidence intervals will be underestimated if the uncertainty caused by the non-response in the second phase isn't taken into account in the variance expression. Comparing the variance expression for the horwitz-thompson estimator with the variance expression when the uncertainty caused by the non-response is taken into account shows this. In this example the sample design is simple random sampling.

The estimator of the population total is;

$$\hat{t}_{\pi} = N\bar{y}_s$$

where

$$\bar{y}_s = \sum_s y_k / n$$

n = Number of enterprises in the sample

The variance expression for this estimator follows below;

$$V_{SI}\left(\hat{t}_{\pi}\right) = N^2 \frac{1-\left(n/N\right)}{n} S_{yU}^2$$

where

$$S_{yU}^2 = \frac{1}{N-1}\sum_u \left(y_k - \bar{y}\right)^2$$

is the variance in the population for the study variable.

One common way to deal with non-response in surveys is to replace the number of enterprises in the sample with the number of responding enterprises in the sample. This will make the variance higher and therefore also the confidence intervals longer. If we use a stratified sample, the number of enterprises in the sample per strata is replaced with the number of responding enterprises in each stratum. This is building on the assumption that the non-response is randomly distributed within strata, which will simplify the variance expression. However, this assumption will often be a rough simplification of the reality. For example within strata, larger firms will probably respond more often than smaller firms. In order to illustrate how the variance is affected by non-response, the variance expression in case of simple random sample with non-response is shown. It is assumed that the non-response is randomly distributed within predefined response groups, h.

$$\hat{t}_{c\pi^*} = \frac{N}{n} \sum_{h=1}^{H_s} n_h \sum_{r_h} \frac{y_k}{m_h}$$

where

m_h = Number or responding enterprises in

each response group

The variance expression will then take the following form;

$$V\left(\hat{t}_{c\pi^*}\right) = V_{SAM} + V_{NR} =$$

$$N^2 \frac{1-(n/N)}{n} S_{yU}^2 + \frac{N^2}{n^2} E_p E_m \left(\sum_{h=1}^{H_s} n_h^2 \frac{1-f_h}{m_h} S_{ysh}^2 / S \right)$$

The first term in this expression, V_{SAM}, is the familiar variance for estimation of a population total with the Horwitz-Thompson estimator in case of simple random sampling of size n. This means that the first component in the expression will take the uncertainty in the estimator caused by the selection of a sample into account, while the second component will take the uncertainty in the estimator caused by the non-response into account. This variance expression is actually exactly equal to the variance expression in case of a two-phase sample where the second phase sample is equal to m_h.

In the case of CVTS3 the sample design is stratified simple random sample. The expressions above will then reflect the sample within a stratum. Each stratum could be divided into response groups, where it is assumed that the enterprises probability to respond is the same. This response groups could for example be a more detailed division according to the size of the enterprise.

References

Wretman, J. (1994). “Estimation in Sample Surveys with Split Questionnaires,” Research Report No. 1994:3, Department of Statistics, Stockholm University.