European Centre for the Development
of Vocational Training

Workshop Continuing Vocational Training Survey: relevance, lessons learnt and ways forward

29 / 30 June 2009, Cedefop

From CVTS3 to CVTS4:

learning from countries' experiences ensures success

Ways forward to improve CVTS
Lessons learnt in the Czech Republic

Josef Kotynek, Czech Statistical Office

At the beginning of my presentation I would like to state the fact, that we realised CVTS in the Czech Republic for the second time. For the first time it was in 2000 year for reference year 1999 as CVTS 2 and for the second time it was in 2006 year for reference year 2005.

I would like to mention, first of all, the procedures or steps, which we gradually realised in order to achieve the best results, in terms of reliability, non-response, comparability, timeliness, punctuality, coherence, accessibility and respondent's burden.

I would split mentioned procedures or realised steps into **2 fundamental categories** :

1. **at administrative and organizational level (typical for preparative phase)**

- as very practical I consider to establish an Interdepartmental Working Group, composed of specialists, dealing with educational statistics and having practical experience with human resources in enterprises (incl. e.g. Labour or Trade unions, Association of industry and transport, etc.) – such WG serves as a continuous advisory, consultative body;
- determine a statistical team, which will be participating in the project at all organizational levels of the statistical office, and at the same time to assign to all of them relevant tasks;
- compile basic instruments for managing of the project, especially then :
  - make up the working schedule, involving all activities of the preparative and implementation phases of the project;
  - calculation of all costs (incl. external institutions, participating in this project), linked with the project from its beginning to the end;
  - compilation of the grant application form (incl. budget and all, required annexes);
  - implement on PC all necessary documents, serving to monitoring of all technical and administrative processes, costs, etc.

2. **at level of factual provisions, leading to fulfilment of all targets (typical for implementation phase)**

- perfect translation of the Eurostat's questionnaire in order to have the highest, clearest understanding from respondents units point of view;
- specification of all concepts and definitions in the sense of an adaptation these concepts and definitions at general level to the specific conditions of national economy and accounting system;
- creation of the flow diagram, which is guiding the respondents to correct direction by completion of relevant sections of the questionnaire,
- production of survey guidelines at national level, coming from survey guidelines of Eurostat – this survey guidelines contains very detailed conditions for :
  a) data collection
  b) data processing and

c) flow of all necessary documents, concerning administrative process, connected with grant's principles and at the same time with involvement of all statistical team at lower level (in regions), participating in the fieldwork (dealing with data collection);

- compilation of a so called Technical Project, which reflects at national level in detail every technical procedures and conditions, leading to successful realisation of the survey (it means from sampling frame, sample, variables, imputation principles and record weighting, data file format and transmission rules up to output format of the tables with collected data, necessary after all with the view of making an analysis).

In order to enrich the outputs, coming from questionnaires, it seems to be very practical to specify several so called **national questions**, which we considered within our Interdepartmental WG as very meaningful for next analysis. To be specific, these are as follow:

1. e-learning training (on-line or off-line);
2. total volume of not paid working time spent on the CVT courses;
3. if the enterprise provided CVT courses for specific group of persons on parental leave;
4. co-operation of the enterprises with schools and universities (this question contained then 4 sub questions) and
5. how high was the portion, rate of employed persons, participating in CVT courses or in "other forms" of CVT, within 3 basic groups of employments in the enterprises (1. leading, managing and technical employees; 2. qualified and helpful employees; 3. unskilled employees – in accordance with ISCO), providing CVT courses or "other forms" of CVT.

And now I try to specify, in addition to above-mentioned special instruments or procedures, leading to the best results of the CVT survey, still **other very important moments, which should be taken into consideration**:

1. to have very good and elastic communication with respondents;
2. to offer to the respondents a possibility to fill in the questionnaire by electronic form;
3. to have perfect co-operation in the scope of national statistical team;
4. to have sophisticated system of controlling linkages;
5. to urge the respondents continuously in order to get the least portion of non-responses.

At the conclusion I would point out very briefly at several moments, on which we should concentrate our attention preferably in the future, before starting with the implementation phase of the CVTS 4 project, in terms of:

1. Questionnaire

- **Section B – CVT activities** – the following seem necessary:
  - to give a explanation, that into this category are pertaining also obligatory special trainings, resulting from rules of law (e.g. training of health and safety at work, fire prevention, drivers, electricians, welders etc.);

- to make sure how to solve, where to include so called e-learning activities; we comprehend this training activity as one of the educational forms, perhaps as the third form (in addition to the 2 basic forms – CVT courses and "other forms"), which can be fading especially with "other forms".

- **Section C – CVT courses** – recommendatory comments :
  - question C8 + C9 – better specification of all costs of CVT courses and CVT contributions/receipts in order to avoid ambiguous understanding;
  - for completeness' sake of all costs connected with CVT courses to take into consideration also the total volume of not paid working time spent on the CVT courses.

- **Section D – Training policy** – advisory comments:
  - question D5 - better specification especially in the case "always" (structured interviews are usually carried out once a year);
  - question D 14 – very questionable first of all from point of view of the explanatory power of these corresponding answers;
- **Section F – Initial vocational training (IVT)** – this section has to be completely reconstructed and the questions to be redefined – this is one of the weakest points of the questionnaire.

2. Manual – there are several cases, where it seems necessary to make certain adjustments or corrections, especially in the parts of the Codebook and 3 Checking rules (conditions for variables).

3. Checking system of Eurostat –we consider very important the problem of the time length between sending our survey outputs to Eurostat and the final period of their approval (this process was taking in the case of the Czech Republic almost 7 months!!!).

Comparison of response rates, respondents and responses

– some evidence from the Norwegian CVTS3

In search of an optimal balance between costs and quality with a mixed-mode/multi-strategy survey design

Bengt Oscar Lagerstrøm
Senior Adviser
Division for sample surveys
Statistics Norway
bol@ssb.no

Kjartan Steffensen
Adviser
Division for education statistics
Statistics Norway
stk@ssb.no

Abstract

The paper addresses the possibilities of optimizing the combination of different survey modes and contact strategies in probability-based sample surveys, with a particular focus on the integration of a web option in the data collection framework. Discussions are based on the results from the Norwegian part of the European Continuing Vocational Training Survey 3 (CVTS3). The Norwegian CVTS3 was conducted using an experimental design that manipulated different combinations of contact strategies and survey modes (D1= postal/web with telephone contact by interviewers to encourage use of the web option, D2= postal/web without telephone contact by interviewers, D3= Face-to-face).

The paper suggests that there is considerable scope for reducing costs without undermining quality, by searching for a more optimal allocation of the sample across different modes/contact strategies in the data collection. The comparison of response rates between modes/strategies indicate a strong effect of using interviewers in an enterprise survey, which especially boosted responding by web questionnaire. However, the paper illustrates that even though the use of a more expensive strategy (e.g. D3 instead of D1, or D1 instead of D2) may be highly recommended for some types of firms, it will not necessarily lead to quality improvements in *all* size groups. Cost-efficiency optimisation of a specific mode/strategy should thus be evaluated separately for firms in different size groups.

Comparison of response rates, respondents and responses – some evidence from the Norwegian CVTS3

In search of an optimal balance between costs and quality with a mixed-mode/multi-strategy survey design

Bengt Oscar Lagerstrøm
Senior Adviser
Division for sample surveys
Statistics Norway
bol@ssb.no

Kjartan Steffensen
Adviser
Division for education statistics
Statistics Norway
stk@ssb.no

Introduction

Since the 1990's survey research has been struggling with declining response rates both international (De Leeuw, E. & De Heer, W. 2002, De Heer, W. 1999) and in Norway (Thomsen 2006). Changing modes of data collection may be an effective way of improving response rates, and using the web as an alternative to other survey modes such as mail or telephone is becoming increasingly accepted (Couper 2000). Mixed-mode approaches have the potential to reduce costs (Dillman 2000) and/or increase response rates (Schaeffer & Dillman 1998). Mixed-mode designs also have potential to reduce non-response bias (e.g. Lagerstrøm 2008, Fowler et al 2002, Salvo & Lobo 1997) if the response propensities of the sample units differ across modes (e.g. some people might be more likely to respond to a telephone survey while others might be more likely to respond when using face-to-face interview).

This paper addresses the possibilities of optimizing the combination of different survey modes and contact strategies in probability-based sample surveys, with a particular focus on the integration of a web option in the data collection framework. Our discussions are based on the results from the Norwegian part of the European Continuing Vocational Training Survey 3 (CVTS3). The Norwegian CVTS3 was conducted using an experimental design that manipulated different combinations of contact strategies and survey modes. Further details about the survey and the experimental design groups are given later in the paper.

A key point of interest in this paper is to examine the outcome of different contact strategies/data collection modes used in the CVTS3 with regards to 1) Increasing response and/or the representation and/or 2) reducing costs. We examine if there are indications in the CVTS3 that certain design(s) could allow us do both at the same time, i.e. be more cost efficient. We define cost efficiency by cost relative to quality measured by response rates and conditional bias in survey estimates.

Thanks to Johan Fosen for his contribution.

The survey

CVTS 3 is an official Eurostat survey on vocational training provided by businesses. The survey was performed using an experimental design that manipulated different combinations of survey modes; web, mail, telephone and face-to-face. The Norwegian CVTS 3 is described in a report by Steffensen and Lagerstrøm (2008).

The Norwegian CVTS3 consisted of a core part common to all participating countries and a pilot part including small firms with 5-9 employees and covering additional industries: Agriculture, fishing and forestry, oil, gas, mining, and quarrying, public administration and defence, education and health and social services.

Both data from the core and pilot part is included in this paper. It is also important to note that the Norwegian CVTS3 was conducted at the local level (firms) and not at the enterprise level as in other participating countries.

The data collection framework used in the Norwegian CVTS3 is somewhat different from the one normally used in business surveys conducted by Statistics Norway. Since July 2004 it has been possible to report all mandatory business surveys (CVTS3 was not a mandatory survey) electronically to Statistics Norway. Two web portals for data collection have been used for such electronic reporting; Idun and Altinn, of which the latter is currently the key channel. The proportion of respondents using the Internet in these mandatory surveys is increasing, from 33 per cent in 2005 to 48 per cent in 2007. The goal is to reach 60 per cent electronic reporting in business surveys in 2010. Further measures include more efficient use of different strategies for contact with respondents, extended response deadline for “e-respondents”, and to provide extensive information on the statistical content to the “e-respondents”.

Since CVTS3 was not a mandatory survey and Idun and Altinn are not designed for mix-mode surveys, we used BLAISE IS instead as a common platform for both the web and the face-to face questionnaire in CVTS3. A server outside Idun and Altinn was used to store the electronically captured data.

As mentioned earlier we used an experimental design in CVTS3, with the objective of studying possible instrument effects due to different modes and examine how interviewers could be used most efficiently in business surveys. The design used in CVTS3 is illustrated in table 1. The table is divided into the three different experimental design groups we used, and the three main steps in our data collection; recruiting, responding and reminding non-respondents after the first contact attempt.

Table 1. The survey design groups

	Initial contact (K)	Response mode (M)	Follow-up treatments (T)
D1	Advance letter telling about the purpose of the survey and the response option; mail back the questionnaire that are provided together with the letter or use the URL and the personal access code that also are included in the letter An interviewer call the respondents and ask if (s)he would respond on the available web questionnaire	Return the paper questionnaire or Visit the URL and complete the web questionnaire	Reminder with questionnaire and URL/personal code
D2	Advance letter telling about the purpose of the survey and the response option; mail back the questionnaire that are provided together with the letter or use the URL and the personal access code that also are included in the letter	Return the paper questionnaire or Visit the URL and complete the web questionnaire	Reminder with questionnaire and URL/personal code
D3	Letter explaining the purpose of the survey and that a interviewer would soon take contact for an appointment	Interviewer calling for appointment for F-2-F	

The D3 face-to-face design is clearly different from the other two design groups. What separates D1 from D2 is that D1 was more focused on the web mode through the use of initial calls prompting the potential respondents to complete the web questionnaire rather than the paper questionnaire. As shown in table 3, the gross sample was not equally distributed between the three design groups. D1 was used for 463 units; D2 was used for 3304 units and D3 for 597 of the units in the gross sample. The large firms (250 + employees) in the gross sample were distributed evenly across the three design groups, whereas the majority of the other firms were included in the D2 design.

Method

Costs

Total cost of the different survey designs consist of fixed cost and variable cost for each of the three set of actions that are describes in table 1 for each design.

Total costs for a certain design are calculated as the sum of fixed costs (c_0, c_{0km}, a_{0km}), costs related to the initial contact (advance letter) based on costs per unit c_{km} and certain data collection costs per respondent r_{km} (responses) for each of $k=1…K$ wave in a survey[1], and $m=1...M$ mode within the k-th wave:

$$C = c_0 + \sum_{k=1}^{K}\sum_{m=1}^{M}\left(c_{0_{km}} + c_{km}\cdot n_{km}\right) + \sum_{k=1}^{K}\sum_{m=1}^{M}\left(a_{0_{km}} + a_{km}\cdot r_{km}\right)$$

For each experimental group in k-th wave, the number of contacted units n_{km} for the survey mode m depends upon the number of respondents, non-respondents and ineligible units from the previous wave in this group. Of course, number of contacted units in each wave, n_k, is the sum of all modes used in this wave, n_{km}. Most typically, however, each wave uses only one mode of contact.

Similarly, after contact mode in wave k, we have in total r_k respondents, which can be the sum of responses obtained with various modes r_{km}. Again, typically, we have only one response mode at a certain wave, but there may be several modes as well. In our experiment, both D1 and D2, have two optional response modes. After the mail invitation, respondents in these two designs could decide by

[1] Wave is defined as an independent attempt in the data collection strategy for a specific survey. Examples of independent attempts are advance letter and different types of follow-up strategies.

themselves if they wanted to respond using the paper questionnaire (and mail it back), or respond by visiting the provided URL and fill in the web questionnaire. In this specific example, the number of responding modes for this wave of contact is M=2.

Most of the variable costs in a survey are mode specific. Including telephone in the data collection strategy gives costs for using interviewers, telephone lines tracking of telephone numbers and interviewing and data entry. In a mail survey we are faced with cost for paper, print, postage and data entry (scanning or manual). We assume that the fixed costs related to development and programming of questionnaires are the same for all three design groups and is therefore neglected in this paper.

The variable cost components in our study are described in details in table 2 for all three design groups by phases or treatment.

Table 2. Variable cost elements by strategy D1, D2 and D3

	D1	D2	D3
Treatment 1	Cost of package components (envelop, cover letter, questionnaire and reply envelop)	Cost of package components (envelop, cover letter, questionnaire and reply envelop)	Cost of package components (envelop, cover letter)
	Cost of assembling the components (insert and sealing)	Cost of assembling the components (insert and sealing)	Cost of assembling the components (insert and sealing)
	Postage	Postage	Postage
	Cost of telephone contact to encourage the use of the available web questionnaire		Cost of contacting and interviewing respondents
Treatment 2	Cost of package components (envelop, cover letter, questionnaire and reply envelop)	Cost of package components (envelop, cover letter, questionnaire and reply envelop)	
	Cost of assembling the components (inserting and sealing)	Cost of assembling the components (inserting and sealing)	
	Postage	Postage	

Quality

To asses the quality outcome of the different experimental designs/strategies, we'll first examine how response rates in different size groups vary across these designs/strategies. It may e.g. be that a particular design is well suited to assure adequate response rates amongst larger firms, whereas another (and perhaps cheaper) design may be just as appropriate for smaller firms. Although using a web rather than paper questionnaire is not a guarantee of heightened quality per se, it may be that the inclusion of automatic checks in a web questionnaire allows for fewer errors than in a paper questionnaire. Our interest in comparing web and paper response rates in the designs D1 and D2 is thus not only a question of costs, but may also be linked to quality.

Low response rates in a survey may prohibit as detailed analyses based on the data as one would hope for. Statistics Norway has e.g. not found it feasible to publish CVTS3 result tables with detailed breakdowns, due to small cell sizes (and thus uncertain estimates) in such detailed tables. Small cell sizes due to high non-response rates may however not be a major quality concern compared to the possibility of non-response bias. In our assessment of quality such non-response bias is thus also considered.

The non-response bias in a certain experimental design group is estimated as a difference between the true value of the target variable P (if we talk about the percentages) and the sample estimate of the

variable *p* obtained by this experimental group. On this basis, the estimate of the mean squared error (MSE) for the variable *p* is calculated as the sum of the variance and the squared bias:

$$MSE(p) = \text{var}(p) + bias(p)^2$$

The ideal measure for comparing the contact strategies (D1-D3) would be the expected cost multiplied by mean squared error (MSE) of a crucial survey variable. However, in the presence of only one survey data set, we can only calculate the cost for *our* dataset. Regarding MSE, it can be decomposed into the variance and the squared bias. The variance can be calculated even with only one dataset. On the contrary, the bias is only possible to find given our special dataset, and when not explicitly otherwise stated, we will by *bias* refer to this conditional bias. The variance vanishes when we consider the situation given the sample, so in the conditional case, the MSE is reduced to the squared conditional bias.

We assume that the bias is due to non-response. The conditional bias can then be fully observed as long as we know both the net sample and gross sample average and is then simply the difference between these two averages. By the nature of non-response, we lack the gross sample average of any of the survey variable. However, sometimes we have for each individual in the gross sample (and the rest of the population) a register variable which by its content *could have been* a survey variable in the survey in question. We denote such a register variable as a *proxy* variable. We could then calculate the bias for the proxy variable, and use this bias as an approximation to the bias of survey variables of similar kind as the proxy.

For the CVTS3-survey, the register variables available have only been the stratification variables, i.e. industry and number of employees. These may not be suitable as proxy for the main survey variables we're interested in. For CVTS3 such main survey variables could e.g. be whether firms offered any type of CVT, if they offered CVT courses or if they offered other forms of CVT to their employees during the reference year. Instead of using a register variable as a proxy we've chosen a somewhat different approach in which the D3 is used as a control group. The estimates for the main survey variables in the D1 and D2 design groups are compared to the estimates in the D3 design group, and the deviations from the D3 estimates *within each size group*[2] is regarded as an approximation of the conditional bias. Using D3 as a control group in such a manner implies the assumption that the estimates based on data collected through the D3 are less prone to errors and a better approximation of the true value of the survey variables than D1 or D2. Such an assumption may be questioned; however there are several arguments in favour of considering D3 as a reasonable control group.
Several studies have addressed non response and quality issues between modes. The response rate depends critically on the kind of survey you do. Of the three basic modes; face to face, telephone and self administrated, face to face usually garners the highest response rate (Steffensen & Lagerstrøm 2008, Groves & Kahn 1979). A Study by Allyson et al (2003) showed that telephone respondents were more likely to satisfice (as evidenced by no-opinion responding, non-differentiation, and acquiescence), to be less cooperative and engaged in the interview, and were more likely to express dissatisfaction with the length of the interview than were face-to-face respondents.

Results

Only one out of three firms in the sample responded. As shown in table 3, small firms respond at a higher degree than the large ones: from 38 per cent amongst the smallest firms (5-9 employees) to 25 per cent amongst the largest firms (250 employees and above). We also find differences in response rates between our three data collection approaches. D3 is the most effective in this sense. With the D3

[2] Since firms in the sample were not randomly assigned to a design group regardless of the size group to which they belonged comparisons of estimates across design groups should only be made separately for each size groups, and not for the total net sample.

strategy almost 40 per cent of the firms responded. D1 resulted in a 36 per cent response rate, whereas D2 (the main strategy in CVTS3) was the least effective strategy in gaining cooperation from respondents; less than one out of three firms responded.

When we break the different strategies down by size groups we find that D1 was more effective on small firms (less than 50 employees) and that D3 was most effective on large firms (more than 250 employees). For firms with 50 to 249 employees we don't find any differences between D1 and D3 in gaining responses. Furthermore, for the largest firms we find that D3 gained more than double the response rate compared to D2. For all size groups, D2 was the least fruitful approach with regards to achieving a high response rate.

Table 3. Interview and non-response by number of employees and design/strategy

	Interview		Non response		Total	
	#	%	#	%	#	%
Total	1 454	33.3	2 910	66.7	4 364	100.0
Employees						
5-9	345	37.9	565	62.1	910	100.0
10-49	679	35.2	1 252	64.8	1 931	100.0
50-249	302	29.9	709	70.1	1 011	100.0
250 or more	128	25.0	384	75.0	512	100.0
Design/strategy						
D1	166	35.9	297	64.1	463	100.0
D2	1 053	31.9	2 251	68.1	3 304	100.0
D3	235	39.4	362	60.6	597	100.0
Design/strategy D1 by employees						
5-9	35	44.3	44	55.7	79	100.0
10-49	48	45.7	57	54.3	105	100.0
50-249	45	39.8	68	60.2	113	100.0
250 or more	38	22.9	128	77.1	166	100.0
Design/strategy D2 by employees						
5-9	261	36.9	447	63.1	708	100.0
10-49	562	33.7	1 105	66.3	1 667	100.0
50-249	201	26.8	550	73.2	751	100.0
250 or more	29	16.3	149	83.7	178	100.0
Design/strategy D3 by employees						
5-9	49	39.8	74	60.2	123	100.0
10-49	69	43.4	90	56.6	159	100.0
50-249	56	38.1	91	61.9	147	100.0
250 or more	61	36.3	107	63.7	168	100.0

Another important aspect of our experimental design was to find an effective way to increase the use of self-reports by the web mode, in order to ensure lower total data collection and data validation costs (Our experiences from CVTS3 do indicate that the web questionnaires had fewer errors etc to be dealt with in the data validation process). Frome table 4 we can see the distribution by response mode for each of the three strategies. The D3 strategy only involves face-to-face interviews, so this design/strategy is omitted here. The table indicates that the strategy D1 is significantly more effective than D2 in getting the respondents to use the web questionnaire. 56 percent of the D1 respondents used the web questionnaire and only 35 per cent of the D2 respondents. In terms of gross respondents this means that 20 percent of the D1 sample responded on the web questionnaire and only 11 percent of the D2 sample. The D1 strategy is thus almost twice a good as the D2 strategy in providing the preferred response mode.

Table 4. Web and paper proportion by design/strategy

	D1	D2	D1 and D2
Web	55.8 %	35.4 %	38.2 %
Paper	44.2 %	64,6 %	61.8 %
N	165	1 048	1 213

Although these results are based on small sample sizes in some of the experimental groups, a few important guidelines for further research are obtained. The comparison of response rates for different strategies shows that their impact varies across firms in different size groups. Furthermore, we find that a strategy targeted at achieving a higher web proportion amongst respondents may be recommended if a higher web response rate is considered desirable.

However, achieving a higher response rate and contacting respondents to assure a higher web percentage may come at a cost. For all three approaches we can summarize the cost components as shown in table 5.

Table 5. Cost components for design/strategy D1, D2 and D3. Cost per unit. Euro

		D1		D2		D3	
		Costs per unit (Euro)	Units	Costs per unit (Euro)	Units	Costs per unit (Euro)	Units
Treatment 1	Contact	1.8 (P&P) + 8.7 (Phone)	463	1.8 (P&P)	3 304	1.8 (P&P) + 21.3 (contact phone)	597
	Response	1.1 (paper) 0.0 (web)	65 66	1.1 (paper) 0.0 (web)	571 212	28.3 (travel) 17.7 (F2F)	235
Treatment 2	Contact	1.4 (P&P)	332	1.4 (P&P)	2 521		
	Response	1.1 (paper) 0.0 (web)	8 26	1.1 (paper) 0.0 (web)	106 159		
Fixed cost due to scanning		990		990			
Total cost		6 396.6		11 211.3			24 601.0
Costs per unit (net sample)		38.5		10.6			104.7
Costs per unit (gross sample)		13.8		3.4			41.2

The unit cost is lowest for the traditional design with a paper survey with a web option. If we compare the cost per interview (net) between the three strategies, we can show that D1 is 3.6 times more expensive per interview than D2, and that D3 is 9.9 times more expensive per interview than D2 and 2.7 times more expensive than D1. Unfortunately we are not able to derive the unit cost per size group. We see that there seems to be a trade-off between increasing response rates and the costs involved when choosing a given strategy or a combination of different strategies. To further evaluate the trade-offs involved when choosing a survey design for CVTS, we have also looked at cost-efficiency for some of the key estimates of CVTS 3: CVT courses, other forms of CVT and any types of CVT provided by firms. In this case we regard strategy D3 to give the "true" estimate of the proportions of training firms. As we've mentioned earlier it is often stated in the literature that the face-to-face strategy gives the "best" data. We assume that the interviewers have larger impact on the motivation for the firms to respond and to respond correctly on the key variables listed above.

In table 6 we present calculations of squared conditional bias (SCB) for designs/strategies by size groups. The squared conditional bias varies between strategies and size groups for all of the three key variables that we look at here. For CVT courses, the smallest SCB is obtained with strategy D2 for firms with 50-249 employees. The largest SCB is obtained by strategy D2 for firms with 10-49 employees. If we look at other forms of CVT we find that D1 has the smallest SCB for firms with 50-249 firms and that the largest SCB is found with strategy D1 for firms with 10-49 employees. When we collapse any type of CVT into one variable, we find the smallest SCB amongst the largest firms in D2 and the highest SCB for the same size group in D1.

These differences in conditional bias across the different strategies by size groups are quite equivocal and may to a certain extent point towards choosing a different set of strategies than if we only look at response rates or costs individually. The implications of the results presented above are discussed somewhat further in the final concluding part of our paper.

Table 6. Conditional bias and cost efficiencies for key CVTS3 variables by size groups

		Unit costs (net sample)	Unit costs (gross sample)	Squared conditional bias	Cost efficiency indicator (net sample)	Cost efficiency indicator (gross sample)
CVT courses						
D1	5-9	38,5	13,8	9,75	375,3	134,5
	10-49	38,5	13,8	84,29	3245,2	1163,2
	50-249	38,5	13,8	186,78	7191,1	2577,6
	250 and above	38,5	13,8	201,35	7751,9	2778,6
D2	5-9	10,6	3,4	133,88	1419,1	455,2
	10-49	10,6	3,4	450,90	4779,6	1533,1
	50-249	10,6	3,4	1,27	13,4	4,3
	250 and above	10,6	3,4	243,69	2583,1	828,5
D3	5-9	104,7	41,2	0,00	0,0	0,0
	10-49	104,7	41,2	0,00	0,0	0,0
	50-249	104,7	41,2	0,00	0,0	0,0
	250 and above	104,7	41,2	0,00	0,0	0,0
Other forms of CVT						
D1	5-9	38,5	13,8	38,89	1497,2	536,7
	10-49	38,5	13,8	254,95	9815,6	3518,3
	50-249	38,5	13,8	0,01	0,4	0,1
	250 and above	38,5	13,8	227,25	8749,1	3136,0
D2	5-9	10,6	3,4	71,15	754,2	241,9
	10-49	10,6	3,4	195,85	2076,0	665,9
	50-249	10,6	3,4	33,31	353,1	113,3
	250 and above	10,6	3,4	0,38	4,0	1,3
D3	5-9	104,7	41,2	0,00	0,0	0,0
	10-49	104,7	41,2	0,00	0,0	0,0
	50-249	104,7	41,2	0,00	0,0	0,0
	250 and above	104,7	41,2	0,00	0,0	0,0
Any types of CVT						
D1	5-9	38,5	13,8	53,90	2075,2	743,8
	10-49	38,5	13,8	91,66	3529,0	1265,0
	50-249	38,5	13,8	47,11	1813,6	650,1
	250 and above	38,5	13,8	211,53	8143,8	2919,1
D2	5-9	10,6	3,4	52,61	557,6	178,9
	10-49	10,6	3,4	157,92	1674,0	536,9
	50-249	10,6	3,4	21,94	232,6	74,6
	250 and above	10,6	3,4	13,44	142,4	45,7
D3	5-9	104,7	41,2	0,00	0,0	0,0
	10-49	104,7	41,2	0,00	0,0	0,0
	50-249	104,7	41,2	0,00	0,0	0,0
	250 and above	104,7	41,2	0,00	0,0	0,0

Discussion

The purpose of this paper was to explore the possibilities of optimizing the combination of different survey modes and contact strategies in probability-based sample surveys, with a particular focus on the integration of a web option in the data collection framework.

The magnitude of the cost elements in our study is specific for the Norwegian CVTS3. These are certainly factors that might vary from survey to survey, or survey organization to survey organization. Hence, what may prove to be the most cost efficient design in a Norwegian CVTS context is not necessarily so in other countries with a different cost structure. The overall cost efficiency of a given data collection approach is also influenced by whether the approach reduces the workload required in the data processing and validation phase. We've only briefly mentioned such costs when discussing the benefits of increasing the use of web questionnaires, but it could also be interesting to include these costs in a further developed optimization model. The cost data presented in this paper nevertheless demonstrate that our different design/strategy groups vary considerably in terms of the costs per unit in the sample (gross or net). Our assessment of quality reveals that the use of a more expensive strategy (e.g. D3 instead of D1, or D1 instead of D2) does not necessarily lead to quality improvements in *all* size groups.

The comparison of response rates between different modes showed a strong effect of using interviewers in an enterprise survey, which especially boosted responding by web questionnaire. The magnitude of the improvement in the response rate was not huge—roughly 4.0 percentage points greater than not using an interviewer in the contact phase of the survey and 7.5 percentage points if the interviewer also conducted the interviews. However, given the low response rate in the survey, there was almost a 20 percent improvement in the response rate between the main strategy (D2) and the face-to-face strategy (D3). From a cost efficiency perspective the face-to-face effort (i.e. D3) should however perhaps be concentrated only on the largest size groups, as response rates amongst smaller firms is just as high, or higher with D1.

When looking at quality in terms of bias, estimates based on the face-to-face interviews (D3) were used as an approximation of the true population value for the selected variables, the. The results show that the D1 and D2 designs/strategies generally produced a substantial bias compared to D3, but there was not a consistent pattern across size groups and the different key variables in favor of either of the two design groups. Interestingly, for some of the size groups the lowest bias (after the face-to-face group) was produced by the D2 design/strategy, which had the lowest response rate. Results thus suggest that interviewers are effectively used for heightening response rates, but in this case they also seem to present a potential source of higher bias for certain variables. The fact that firms were not randomly assigned to a design group regardless of the size group to which they belonged limits the way in which the quality of the different designs can be compared across size groups. Furthermore the number of responding firms in some "size group x design group"- cells is quite small. One should thus be cautious not to conclude too strongly based on these findings.

Regardless of the somewhat ambiguous patterns identified by our limited analysis of bias, the material presented in this paper suggests that there is considerable scope for reducing costs without undermining quality, by searching for a more optimal allocation of the sample across different modes/contact strategies in the data collection. It is shown that the optimization of a specific mode/strategy should be evaluated separately for each specific size group. The optimality is strongly determined by the bias (which might differ significantly between variables), initial sample sizes and the structure of costs. The mode/strategy performing optimally in one specific situation might not be optimal when, for example, an initial sample size is different. Optimization models and their further developments could allow for analysis of different scenarios with varying initial sample sizes and cost restrictions, thus presenting the basis for successful approaches to the data collection (perhaps with a further strengthening of web options) in a survey like CVTS4.

References

Allyson l., Holbrook, M. Green, C. and Krosnick, J. A. (2003) "Telephone versus Face-to-Face Interviewing of National Probability Samples with Long Questionnaires: Comparisons of Respondent Satisficing and Social Desirability Response Bias" in *Public Opinion Quarterly,* 67:79-125.

Couper, M. P. (2000) "Web Surveys: A Review of Issues and Approaches" in *Public Opinion Quarterly,* 64:464-494.

De Heer, W. (1999) "International Response Trends, Results of an International Survey" in *Journal of Official Statistics,* 15:129-142.

De Leeuw, E. and De Heer, W. (2002) "Trends in Household Survey Nonresponse: A longitudinal and International Comparison" in Groves, R, Dillmann, D.A.,Eltinge, J.L., and Little, R. J. A. (eds.) *Survey Nonresponse*. New York: Wiley.

Dillman, D. A. (2000) *Mail and Internet Surveys: The Tailored Design Method*. 2nd edn. New York: Wiley.

Fowler, F. J. Jr, Gallagher, P. M., Stringfellow, V. L., Zaslavsky, A. M., Thompson, J. W. & Cleary, P. D. (2002) "Using telephone interviews to reduce nonresponse bias to mail surveys of health plan members" in *Medical Care,* 40:190-200.

Groves, R. M., & Kahn, R. L. (1979) *Surveys by telephone: A national comparison with personal interviews*. New York: Academic Press.

Lagerstrøm, B. O. (2008) *Cost efficiency in a mixed-mode survey Evidence from the Norwegian Rent Marked Survey*. Paper presented at the 19th International Household Survey Non-response Workshop. Ljubljana. Slovenia.

Salvo, J. J. & Lobo, A. P. (1997) *The American Community Survey: nonresponse follow-up in the Rockland County test site*. Paper presented to the annual meeting of the American Statistical Association, Anaheim, California, August 1997.

Schaefer, D. R. & Dillman, D.A. (1998) "Development of a Standard E-mail Methodology" in *Public Opinion Quarterly,* 62:378-397.

Steffensen, K. & Lagerstrøm, B. O. (2008): *Continuing Vocational Training Survey (CVTS3): Quality report for Norway*, Documents 2008/10, Statistics Norway.

Thomsen I, Kleven, Ø., Wang, J. H. & Zhang L-C. (2006) *Coping with decreasing response rates in Statistics Norway. Recommended practice for reducing the effect of nonresponse*. Reports 2006/29. Statistics Norway.

Statistics Norway

Comparison of response rates, respondents and responses – some evidence from the Norwegian CVTS3

In search of an optimal balance between costs and quality with a mixed-mode/multi-strategy survey design

Statistics Norway

CVTS3 in Norway
Facts about the survey

- Firms, not enterprises, were sampled in the Norwegain CVTS3
- The scope of the target population:
  - Core part common to all participating countries (20 NACE x 3 size groups)
  - Additional units surveyed:
    - Firms with 5-9 employees
    - Firms in Agriculture, fishing and forestry, Oil, gas, mining, and quarrying, Public administration and defence, Education and Health and social services.
- Voluntary survey
- Different data collection platform used than in other mandatory business surveys
- Three different approaches used in the experimental design for CVTS3

Statistics Norway

	Initial contact (K)	Response mode (M)	Follow-up treatments (T)
D1	Advance letter telling about the purpose of the survey and the response option; mail back the questionnaire that are provided together with the letter or use the URL and the personal access code that also are included in the letter An interviewer call the respondents and ask if they would respond on the available web questionnaire	Return the paper questionnaire Or Visit the URL and complete the web questionnaire	Reminder with questionnaire and URL/personal code
D2	Advance letter telling about the purpose of the survey and the response option; mail back the questionnaire that are provided together with the letter or use the URL and the personal access code that also are included in the letter	Return the paper questionnaire Or Visit the URL and complete the web questionnaire	Reminder with questionnaire and URL/personal code
D3	Letter explaining the purpose of the survey and that a interviewer would soon take contact for an appointment	Interviewer calling for appointment for F-2-F	

3

Statistics Norway

Variable cost elements by design/strategy

	D1	D2	D3
Treatment 1	Cost of package components (envelop, cover letter, questionnaire and reply envelop) Cost of assembling the components (inserting and sealing) Postage Cost of telephone contact to encourage use of the available web questionnaire	Cost of package components (envelop, cover letter, questionnaire and reply envelop) Cost of assembling the components (inserting and sealing) Postage	Cost of package components (envelop, cover letter, questionnaire and reply envelop) Cost of assembling the components (inserting and sealing) Postage Cost of contacting and interviewing respondents
Treatment 2	Cost of package components (envelop, cover letter, questionnaire and reply envelop) Cost of assembling the components (inserting and sealing) Postage	Cost of package components (envelop, cover letter, questionnaire and reply envelop) Cost of assembling the components (inserting and sealing) Postage	

4

Statistics Norway

Cost measures

$$C = c_0 + \sum_{k=1}^{K}\sum_{m=1}^{M}\left(c_{0_{km}} + c_{km} \cdot n_{km}\right) + \sum_{k=1}^{K}\sum_{m=1}^{M}\left(a_{0_{km}} + a_{km} \cdot r_{km}\right)$$

5

Statistics Norway

Quality measures

- Is the design/strategy successfull in generating an acceptable response rate amongst firms in the different size groups?
- Is the design/strategy successfull in encouraging the use of web rather than paper questionnaire.
  - The inclusion of automatic controls and filters could plausibly lead less errors in the web- than in the paper questionnaires.
    - Our experiences from the CVTS3 data validation supports this notion
- Is the design/strategy successfull in avoiding biased estimates in the different size groups?
  - Bias is measured in a fairly crude manner as conditional bias, using D3 as a control group thought to provide a better approximation of the true value of the survey variables than D1 or D2.

6

Statistics Norway

Results: Costs (€)

	D1	D2	D3
	Costs per unit	Costs per unit	Costs per unit
Costs per interview	38,5	10,6	104,7
Costs per gross sample unit	13,8	3,4	41,2

7

Statistics Norway

Results: Response rates

	Response		Non-response	
	#	%	#	%
D1 by emp				
5-9	35	44,3	44	55,7
10-49	48	45,7	57	54,3
50-249	45	39,8	68	60,2
250+	38	22,9	128	77,1
D2 by emp				
5-9	261	36,9	447	63,1
10-49	562	33,7	1 105	66,3
50-249	201	26,8	550	73,2
250+	29	16,3	149	83,7
D3 by emp				
5-9	49	39,8	74	60,2
10-49	69	43,4	90	56,6
50-249	56	38,1	91	61,9
250+	61	36,3	107	63,7

8

Statistics Norway

Results: The use of web questionnaire

	D1	D2	D1 and D2
Web	55,8 %	35,4 %	38,2 %
Paper	44,2 %	64,6 %	61,8 %
N	165	1 048	1 213

9

Statistics Norway

Results: Conditional bias in key estimates

	Descrip. of key estimate	*Percentage of firms offering CVT courses*		*Percentage of firms offering other forms of CVT*		*Percentage of firms offering any types of CVT*	
	Unit costs (net sample)	Squared conditional bias	Cost efficiency indicator (net sample)	Squared conditional bias	Cost efficiency indicator (net sample)	Squared conditional bias	Cost efficiency indicator (net sample)
D1 by emp							
5-9	38,5	9,75	375,3	38,89	1497,2	53,90	2075,2
10-49	38,5	84,29	3245,2	254,95	9815,6	91,66	3529,0
50-249	38,5	186,78	7191,1	0,01	0,4	47,11	1813,6
250+	38,5	201,35	7751,9	227,25	8749,1	211,53	8143,8
D2 by emp							
5-9	10,6	133,88	1419,1	71,15	754,2	52,61	557,6
10-49	10,6	450,90	4779,6	195,85	2076,0	157,92	1674,0
50-249	10,6	1,27	13,4	33,31	353,1	21,94	232,6
250+	10,6	243,69	2583,1	0,38	4,0	13,44	142,4

Conditional bias is calculated with the assumption that the estimates from units in D3 provide the best approximation of the true proportions of training firms.

10

Statistics Norway

Discussion and conclusions

- Results indicate that cost and quality concerns may be balanced in a better way if we systematically take into consideration that *the optimal design/strategy varies between firms of different size*.
- The use of interviewers, in the contact phase (D1) or in the interview (D3) has a positive effect on response rates.
  - From a cost efficiency perspective the face-to-face effort (i.e. D3) should perhaps be concentrated only on the largest size group.
- The use of interviewers is successful in encouraging answers by web questionnaire.
  - Extra costs of telephone contact outweighed by error reduction and reduced costs in the data validation process?

Statistics Norway

Discussion and conclusions

- The quality-cost balance is somewhat different when we look at conditional bias instead of response rates.
  - D2 generated the lowest response rates, but for some size groups the conditional bias when using D2 is lower than in D1.
  - Interviewers as a possible source of higher bias for certain variables?
    Not clear why.
    - Bias may vary considerably between variables.
    - The number of respondents in D1 is low.
    - More elaborate measures of bias may be needed to draw clear conslusions.
- Optimality-considerations are strongly influenced by bias, inital sample sizes and the structure of costs.
  - E.g. with a limited inital sample size the need to stress a sufficent response rate may be particularly important to allow for estimations to be made with a sufficiently detailed breakdown of firms.

Do small enterprises spend less time to respond to the questionnaire?

Jeroen Nieuweboer, Statistics Netherlands

In the Netherlands, CVTS2 was carried out as a face-to-face interview survey. With the CVTS3, the aim of Statistic Netherlands (SN) was to dispatch questionnaires and these questionnaires should mainly be electronic. In our case, this means dispatching the questionnaires by e-mail, not a web-based survey. All enterprises in the survey received a CD-ROM with software which enables the enterprise to read, fill in and return the questionnaire send to it by e-mail. This method of data collection is used by business statistics (production). So enterprises should be familiar with it.

From a population of 60,050 enterprises, we drew a sample according to the conditions of Eurostat. The sample consisted of 5,837 enterprises. Enterprises with over 250 employees were all included in the sample.

From the sample, we picked 40 enterprises that were to take part in a pilot. The purpose of the pilot was twofold: testing the questions and definitions on 'comprehensibility' and testing the user-friendliness of the electronic questionnaire.
Encountered difficulties: Definition problems and lack of clarity when filling-in the questionnaire.

A month before the initial dispatch of the questionnaire we sent a letter to all enterprises in which we announced the survey and asked the enterprises to give us a correct email address. The letter was accompanied by the mentioned CD-ROMS.

The distribution between electronic and paper questionnaires was 60 % electronic and 40 % paper at time of the initial dispatch.

The first reminder was both electronic and paper. The second reminder was only in paper and the third reminder was by telephone.

Eventually we got a response of 76,7 %. The distribution between received electronic and paper questionnaires is 36 % and 64 %. This is quite a shift towards the traditional way of filling in a questionnaire.

Main problems:
- Installation software (CD ROM)
- E-mail address: does this address still exist?
- E-mail address: is this the right person?
- SPAM filter blocked the questionnaire
- More people in an enterprise are needed to fill in questionnaire

Looking at the quality of the received data, we concluded that 2,5 % of paper response was not useful at all, but the electronic responses were all useful except one. Furthermore, the editing (re-contacting) and imputation needed on key and other variables was much higher for the paper questionnaire.

Looking at sections A and C of the questionnaire, we see that enterprises have difficulties in providing detailed information (gender, age) on training in the enterprise. When the size of

the enterprise increases, the data quality decreases. This is true for both the electronic and paper questionnaire. But with the electronic questionnaire the correlation is much lower.

The average time spent on the questionnaire by enterprises with CVT courses is 158 minutes. The average of the electronic questionnaire is 18 minutes higher but this difference is not significant. However, according to our data, we find that smaller enterprises spend less time on the questionnaire.
On the one hand this seems logical, because a smaller workforce could mean a better (easier) on training activities. But on the other hand, one can expect bigger enterprises to have a more thorough policy towards education (including a specific budget), so this is better documented and therefore easier to provide data on training.

Table 1 – Average time spend on questionnaire (minutes)

	Total	**Electronic**	**Paper**
10-19 employees	98	94	99
20-49 employees	110	101	116
50-249 employees	146	167	131
250-499 employees	199	186	211
500-999 employees	235	219	252
>1000 employees	316	379	273
Total	**158**	**169**	**151**

We can conclude that bigger enterprises need more time to fill in the questionnaire and they provide data of poorer quality. An electronic questionnaire improves the quality of the received data. It seems to force enterprises to at least give their own estimations, which we think is better than imputing. Because there doesn't seem to be any difference in time spent on electronic and paper questionnaires, electronic questionnaires should be preferred (from the point of view of NSI!). On the other hand, enterprises seem to prefer paper questionnaire over electronic ones. But the experienced technical problems with questionnaires by e-mail most likely forced enterprises to change to the paper version.
The technical problems could probably be avoided by using a web-based survey.

Do small enterprises spend less time to respond to the questionnaire?

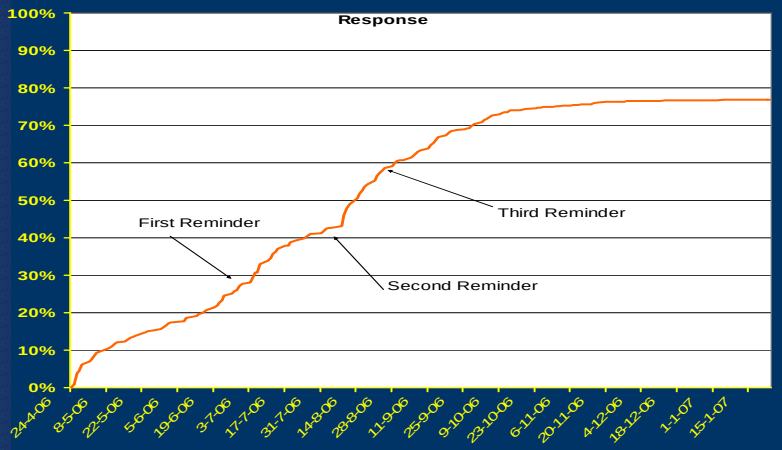
Jeroen Nieuweboer
Statistics Netherlands (CBS)
Cedefop CVTS Workshop, 30 June 2009, Thessaloniki

Content

- Methods of data collection
- Response and quality per method
- Time spent per method and size enterpr.
- Conclusion

Initial sending and response

Population	62 050
Sample	5 837 (>250 empl. included)
Response	4 476 (76,7%)



Two MODES of Data collection

1. Paper questionnaire (postal)

2. Electronic questionnaire (not internet!)

-Data collection used by business statistics (install software (CD ROM), receive e-mail with questionnaire, send back by e-mail)
- Lower burden for the enterprises (assumed)

PILOT: 40 enterprises → Aim: improve electronic questionnaire

Problems:
- Installation software (CD ROM)
- e-mailadress: does this adress still exist?
- e-mailadress: is this the right person?
- SPAM filter blocked the questionnaire
- more people in enterprise needed to fill in questionnaire

Dispatch and Response

	Initial dispatch	Initial disp resp	response
Electronic	59%	60%	36%
Postal	41%	40%	64%
Total sample	5837		
Total response		4476	4476

Shift from electronic to paper questionnaire

Per size class

%	send	response	diffrnce
10-19 employees	16	21	5
20-49 employees	54	36	-18
50-249 employees	75	39	-36
250-499 employees	81	44	-36
500-999 employees	81	43	-38
>1000 employees	80	40	-40
Total	60	36	-24

Data Quality (1)

(Enterprises with CVT courses)

Paper → 2,5% not usefull at all

Electronic → all usefull except 1

Re-contacting/Editing/Imputation needed on key and other variables (Sections A and C)

Electronic average 5,8% (of questionnaires per variable)

Paper average 16,3% !

Example: Paid working time spend on CVT

	Electronic	Paper
C3tot	7,0%	15,9%
C3i/e	7,1%	17,4%
C4m/f	7,0%	23,5%

Data Quality (2)

Enterprise size and data quality seem to correlate

Bigger enterprises

→ poorer quality (more editing/imputing needed)

Correlation much higher with paper questionnaire

Time spent on questionnaire (1)

(Enterprises with CVT courses)

In minutes (averages)	Total	Electronic	Paper
10-19 employees	98	94	99
20-49 employees	110	101	116
50-249 employees	146	167	131
250-499 employees	199	186	211
500-999 employees	235	219	252
>1000 employees	316	379	273
Total	158	169	151

Time spent on questionnaire (1)

(Enterprises with CVT courses)

In minutes (averages)	Total	Electronic	Paper
10-19 employees	98	94	99
20-49 employees	110	101	116
50-249 employees	146	167	131
250-499 employees	199	186	211
500-999 employees	235	219	252
>1000 employees	316	379	273
Total	158	169	151

Time spend on questionnaire (2)

Do small enterprises spend less time to respond to the questionairre?

or…

Do big enterprises spend more time to respond to the questionairre?

YES, they do!

CONCLUSIONS

Results are based just on quick analyses!

Bigger Enterprises

- → more time spent on questionnaire
- → and poorer data quality

"Electronic mode"

- → negative: more time spent? No proof
- → positive: better data quality
- → correlation size and time spend remains (lesser extent)
- → forces enterprise to make own estimations (better then imputation?)

CONCLUSIONS

The bigger the enterprise the more difficult to fill in the CVTS(3) questionnaire

→ reconsider variables questionnaire? (male/female, age, hours spend)

(→ is web based survey the future?)

Enterprises – the decisive point in the data collection process

Hans-Joachim Blömeke – Federation of German Paper Industries Employers' Associations

CVT is of great importance for the German Paper Industries (view chart):

Age Group			
Adolescent	Low birth rates	Lack of education in MINT academics	Insufficient number of trainees
Senior colleagues	Start of old age pension will be changed from the age of **65** to the age of **67**!!!	Skill enhancement is urgently needed to keep them up to the state of technical progress	To keep them in job the fading physical strength has to be compensated by growing mind strength

Papierzentrum Gernsbach (Paper and educational center in Gernsbach / Black Forest)

For the German and Swiss Paper industries, the Papierzentrum Gernsbach is the central institute for vocational, advanced and further education.

With 600 beds, it is booked for approx. 95.000 overnight stays per annum.

There are various CVT courses offered, concerning technique, papermaking, chemists, languages, human resources, management, consultancy, computer, health and safety, fit for studies, distance learning, etc.

Statistics

1. 2389 participants in training courses (excl. vocational training) held in the Papierzentrum Gernsbach in 2008;

2. 495 participants in training courses held by instructors of the Papierzentrum Gernsbach in the companies in 2008;

3. Unfortunately, we are unable to give a number of the training courses held in the companies with other than our instructors.

   Here the questionnaire could give answers.

Problems

The Problem with statistics in Germany is that there are plenty of questionnaires that have to be lawfully filled in by the companies.

We know about **22** statistics, which has to be served monthly / half a year or yearly. They are collected by associations, regional authorities and the Federal bureau or the Federal Statistical Office.

Especially small and medium-sized enterprises do not have enough employees at their disposal to handle the flood of paper.

EU-Questionnaires are also mostly written in English or French and only large enterprises have personnel qualified in foreign languages.

Solutions:

If there are further questionnaires, they should be:

- Ii the national language;
- short (only the essential questions / understandable question which do **not** need long explanations);
- theme of questions should be interesting for the enterprises;
- results should be handed out to the enterprises / no Internet link.

Enterprises – the decisive point in the data collection process

Statement by Werner Auracher – Austrian paper industry

General remarks:

CVT is of great importance for the paper industry

Producing paper requires high technological knowledge

Due to the rapid technological progress lifelong learning is a key issue for our employees

A qualified skilled workforce is an essential asset on international markets

The Austrian paper industry runs its own training center (Ausbildungszentrum Steyrermühl http://www.papiermacherschule.at/), offering a wide range of training for employees in paper mills and paper related sectors (suppliers, converters etc.)

Statistical comparisons on national or international level:

So far we have not been very successful in getting comparable figures regarding training & education

Definition of training is different from mill to mill both on national and international level

International groups have their own comparisons and figures – not comparable!

Different structures of CVT in European countries (f.e. Austria and Germany have own training institutions for paper industry, in other countries CVT organized on mill level; training on the job, etc.)

Significance of the CVT training survey of Eurostat:

Among our members both Cedefop and the survey and its results are completely unknown

In our opinion the results are not interesting for our mills, because it does not apply exactly to the paper producing sector and is mixed up with converting, printing and other sectors, which have completely different training structures

No direct, but indirect benefits for mills:

The survey provides a good overview and useful information for training policy stakeholders and politicians on national & European level

Useful tool for preparing political decisions

More advertising on national level: role of Cedefop, significance of survey, benefits etc.

Assessment of Quality in the Swedish CVTS3[1]

Executive summary

Lennart Forssén, Statistics Sweden, Population and Welfare Department

Reflexions for the future

Survey specific reflexions for the future

We should start with the planning of future surveys already at the end of 2007 or in the beginning of 2008. The start of this planning should of course be in the form of a final CVTS3 WG-meeting before all CVTS3 WG-representatives have dispersed. In this meeting countries will be given the opportunity to describe the situation and what they think about the future for the survey when still having experiences from their CVTS3 in relatively fresh memory. A contract group consisting of representatives from several countries should be given the main responsibility for looking into the question on what kind of information is needed to be collected and how to do that in the best way. To that contract group may perhaps also be tied one or several minor groups looking into certain limited issues. Several, preferably all, union countries should also be involved either by participating in some contract group, a Task Force group, a reference group or in some other way more or less from the start of the planning.

One important reason for starting the planning of future surveys already at the end of 2007 is that in that way the planning situation for both countries and Eurostat will be much better than in CVTS3. For countries it would then perhaps be possible to forewarn the enterprises or their organizations about the survey, its purpose and the kind of questions that will be included in the survey. Several enterprises informed us both in CVTS3 and in connection to our CVTS2 that they would have been in a better position to answer our question had they been informed a year in advance.

The data quality in the Swedish CVTS3 would probably have been a lot better had we performed face-to-face interviews with the biggest enterprises. However, the costs for doing that would have been extremely high. A rough estimate indicates that the total cost for the Swedish CVTS3 in that case, with every thing else unchanged, would have been almost 50% higher.

Due to the globalization we can expect that we will get more problems with CVT-activities taking place abroad. It will probably be relatively important already in connection to CVTS4 to see to it that this kind of activities are measured and also that we do not risk double or triple counting instead of missing data.

The first version of the European checking tool in CVTS3 was not available until the 27th of March 2007. Many of the tests limits in the common checking tool were also too tight which made the use of this instrument unnecessarily difficult. At the moment when the checking tool became available most countries had probably since quite a long time ended all their re-

[1] For further information, please contact Mr Anders Ljungberg, anders.ljungberg@scb.se, at Statistics Sweden. Mr Forssén is now retired from Statistics Sweden.

contacts to the enterprises. In other words the checking reasonable to do at that moment was more or less only to guarantee that correct codes had been used or that no serious errors still existed in the files that had to be delivered to Eurostat. Had the checking tool been available one year earlier and had not the test limits been quite as rigorous the situation both for Eurostat and in CVTS3 participating countries would have been much better. Many countries would then probably have been able to deliver their micro data files to Eurostat much earlier than they did in CVTS3.

The checking tool should also be improved in many different ways before being used in the next survey. In the present version it is for instance too difficult to change test boundaries for the user. Quite often it is not possible to see the whole logic behind some of the tests now without having to scroll sideways. In fact, some of the test expressions were so long that it was even difficult to see the expression at all. A sideways scrolling list is lacking. The logic behind the selection of units going into each test could perhaps be shown in a separate line. In that way some of the test expressions would be shorter and could perhaps more often directly be shown on the screen without the user having to scroll sideways. The checking tool should be one of the items on the agenda for discussion at future CVTS WG-meeting.

In some of the tests in the common testing tool enterprises with low costs per employee, per course participant or per course hour were marked as possible errors. With these kind of tests it is usually quite difficult to know whether we have received wrong information or not from our respondents. The reasons are many. For instance a substantial part of the course hours may quite often be internal courses more or less in the form of lectures with many participants (perhaps the whole staff) and extremely few teachers. Some providers also quite often do not charge anything or only minor sums for their involvement in courses within different enterprises. Such providers are especially frequent among trade unions, employer's associations, chambers of commerce, sector bodies and for enterprises whose main activity is not training (e.g. equipment suppliers). However, in these cases we sometimes risk really serious errors here when an enterprise perhaps may have provided us with costs for something else than course costs or costs in another currency than asked for or perhaps even costs in thousands instead of in let's say crowns, forints, korunas etc. Could, in the future, these kinds of tests be improved by for instance involving other variables in the testing we could perhaps better be able to catch units where we have received wrong information. In CVTS3 most of these signaled and possible errors were units were we simply had received correct information from the enterprises.

The number of WG-meetings have since quite some time now been heavily reduced. One consequence of that is that the need for different countries to better be able to follow the work progress with the survey in other countries has increased substantially. Many countries would probably be helped a lot if Eurostat at regular time intervals reported back to all countries the problems that have come to its knowledge and how these problems have been solved by other countries. In CVTS3 countries were many times working more or less alone and therefore had to come up with their own solutions to different problems. Had there been regular discussions between different countries and Eurostat many countries would probably have saved a lot of time and costs and may also have been able to better solve some of their problems.

Quite often during the work with our CVTS3 we felt the need to contact and discuss certain issues with one or a couple of other countries that we believed to have similar conditions. In order to make it somewhat easier to do that it is necessary that Eurostat in the future *always provide countries with participating lists from all WG-meetings and also with current lists of contact persons responsible for CVTS-issues in different countries*. These lists should contain

information about name, postal address, e-mail address, office telephone number, mobile number and fax number.

Another way of improving discussions and forwarding good solutions and best practices to other countries could perhaps be in the form of an electronic discussion group involving at least Eurostat and all the participating countries CVTS WG-delegates.

The Control table values for Ref 35, 36, 58 and 59 and other similar table values in future surveys with normally quite small table values should be delivered with at least one decimal. To check values that usually for instance are in the order of let's say 1-5 it is imperative to use better precision than whole numbers without any decimals.

The staff resources put aside for CVTS both by countries and by Eurostat will probably have to be increased in order to fulfill some of the proposals mentioned earlier in this paragraph. Countries will also have to be obliged to report back to Eurostat as quickly as possible when they observe problems of possible interest for other countries or when they change the person responsible for their CVTS.

General reflexions for the future

Increased international competition due to the globalization and quick technical developments can be expected in the foreseeable future. In order for the EU countries to be able to be competitive on the global market it will be necessary to keep up the pace with other countries and economical regions when it comes to technical developments and also by continuously performing different tasks in a better and smarter way. However, in order to be able to do that one has also to see to that the competence levels in the labour force if possible and preferably are higher than in competitive countries and regions. In the foreseeable future we will have to compete with a lot of countries with much lower labour costs. It will take many decades before equilibrium can be reach when it comes to differences in wages and salaries between different countries in the world. We cannot only rely on the fact that with increased productivity later on with a time delay seems to follow higher wages and salaries in Newly Industrialized Countries (NIC) according to earlier experiences of the development in for instance countries like South Korea, Hong Kong, Taiwan and Singapore. To compete with the much lower labour costs in the third World countries we simply will have to perform different work assignments smarter. The public education system in a country plays an important role in order to be able to do that. One should also take into account that different parts of the public education system are interconnected. If there are deficiencies at some of the lower levels in the public education system in a country, that in itself will of course lead to worse results even at higher education levels within the system. Even though different parts of the public education system play an important role in promoting competitiveness in a country one should not forget the importance of promoting CVT-activities in the labour market. CVTS can and should play an important role in describing development within this area. However, in order to be able to do that, major changes in the survey and its performance will have to be done. Some of these changes will involve a lot of pre-planning that will be time-consuming and costly. To perform future surveys will also most probably be much costlier than today. The rest of the text in this paragraph contains a number of descriptions, assumptions and suggestions both concerning the situation today and in the foreseeable future.

The kind and volume of different CVT-activities in an enterprise quite often depend more on how an enterprise is organised than on what products or services are produced, which branch of Industry or size group the enterprise belongs to. The need of developing some kind of new

classification systems of enterprises other than NACE- and size group is therefore of crucial importance and will certainly be even more important in the future when describing what is taking place within different enterprises and on the labour market.

Many researches and think-tanks in the private sector today mean that since around 1985 a new paradigm has come into existence and will for each year more and more influence enterprises, organisations and whole societies. The main reason for that is spelled *Globalization*. In the old paradigm most of the trade crossing borders were in the form of goods or products. Since around 1985 more and more of the trade crossing borders are in the form of tasks and services.

The first wave in the new paradigm had in some way usually quite strong connections with the ICT-sector. The reason for that had often to do with the fact that tasks in the first wave usually did not need face-to-face contacts between the buyer/consumer and the seller/provider. Input and or output were also quite often in the form of direct ICT-transferable information. Call centre activities are example of tasks traded across borders in the first wave.

In the second wave, that has started quite recently, many researchers and think-tanks mean that in principle almost any tasks will be tradable across borders. Some work assignments of course will still be difficult to transfer across borders even in the future. It will perhaps also arise strong resistance against the globalization and possibly also a tendency towards isolation in some countries. Cultural differences and lack or deficiencies in infrastructure will of course also be example of obstacles for trade in tasks in the foreseeable future. However, we can for several decades now foresee extreme differences in economic compensation for labour between different countries that often are not reflected in differences in education levels or skills in the work force in different countries. Factors like that will by and large guarantee that trade in tasks will increase for quite some time in the future.

Many different enterprises have in some way need of similar functions within their companies. Usually and in most enterprises today there exists some kind of departments/units/functions for economical issues, juridical issues, ICT-issues, advertising and marketing, purchasing, selling and/or trading etc. Some think-tanks today mean that that what's just mentioned are something that will simplify making tasks transferable across borders in the second wave of the Globalization.

Whether a task will be tradable or not will sometimes change more or less over a night. Due to the fact that both employees and enterprises often will be affected differently and quite often also in unforeseeable ways will increase the uncertainties for all. Differences in economic return of labour will for the same reason probably also increase. Winners (both enterprises and individuals) will get bigger markets and their services and products will be more often required. Losers on the other hand will quite often go bankrupt or lose their jobs. In order to survive, enterprises will have to decentralise a lot of tasks to local units and at the same time they will also probably have to centralise some other tasks. Outsourcing and offshoring of tasks will increase. Due not only to the ICT-development but also to the development of more and more complicated robots we can expect that more and more work assignments not necessarily have to be performed in the same location as the actor. *In other words it will in the future be much more important to know what is taking place within different enterprises both at central and local level. It will in other words in CVTS4 or in similar surveys not be enough just to ask the enterprises at central level about CVT-activities.*

In the future it will be necessary to use other background variables than we use today when describing the development in enterprises. An increasing part of the products/services/tasks an enterprise are dealing with will in the future more often than today be provided or produced by other companies or in other countries. *The difficulties to in advance forecast where the winners respectively the losers will be, will increase. It will not be enough to produce and present statistics for enterprises by NACE- and/or size groups only. We simply will have to develop new ways of classifying enterprises in order to better describe what is happening within different types of enterprises.* These new kinds of classifications of enterprises will be needed, not only in CVTS, both when allocating samples and when presenting statistics in order to better be able to describe what is going on within the enterprises. One important aspect in this connection will probably be to describe how enterprises are organized and how different enterprises have solved co-ordination problems between different levels in the organization. Major structural changes within and sometimes also between many different Branches of Economy are expected in the future. Productivity seems to be highly correlated to the level of international contacts that an enterprise has. Spin-off effects of co-operation with other enterprises in other areas both nationally and internationally are important and will probably be that even more in the future. The importance of international co-operation will probably also make it sometimes and to some extent somewhat irrelevant to talk about French, German, Swedish etc companies. Instead it will probably be more and more important to talk about different company cultures. Global and regional markets are considered to play a more important role even for quite small enterprises in the future. As a consequence of that, it will be natural with more co-operations across borders and with enterprises in other sectors on the labour market. It will be vital to reflect things like that in future enterprise statistics. For the CVTS it will for instance also be important to know how different CVT-activities are organized, supported and decided on at different levels within different types of enterprises.

What has been mentioned above in this section of the document may perhaps reflect why Sweden in CVTS3 has observed more of decentralisation of CVT-activities in many enterprises than in CVTS2. Perhaps that is also the reason why some enterprises indicated or said that they had a feeling that CVTS3 did not always describe what was going on today within their enterprise and in the economic life in general. “Enterprise statistics quite often today do not cover what we consider as most important”, one respondent said. Sometimes our respondents said that they used other concepts in order to better describe the education and training activities that took place within their enterprise.

Partly due to what has been mentioned above, in the future, the ideal state should be to put questions both to the enterprises at central and at local level. Both these surveys should preferably also be complemented by statistics from a survey to the employees in the selected enterprises. This proposal may perhaps seem somewhat unrealistic for several reasons. It all to a great extent depends of course on how important one considers it to be to be able to connect information for the different surveys to the same enterprises. Increased survey costs will naturally also be an almost insurmountable obstacle. However, to my opinion we will simply have to move into the above proposed direction in the future, if we still will be able to provide statistics describing what is taking place in different enterprises, in the economic life in general and in different countries. However, in order to be a little more realistic we will probably have to drop the necessity of directing a survey to within the selected enterprises and local units employed personnel. A separate and independent survey to employed individuals will probably have to suffice. We will therefore probably also have to accept that results from such a survey only can be combined to our enterprise surveys results at macro level by variables like NACE- and size group and other similar background variables.

Due to the expected and increased uncertainties concerning which tasks, individuals and enterprises will be losers or winners in the future, it will be of crucial importance for both the society and the enterprises to look into issues of security and social welfare for employees. A separate survey directed to employees should be able to include some such questions. A survey to individuals quite often also provide the most accurate information about volumes of education and training as it is the individuals themselves that participate in different CVT-activities. It will, perhaps for several reasons, not be possible to have both a survey to the enterprises at central level, at local unit level and also to the employees. However, at least two separate surveys to enterprises should be both possible and necessary to go for.

If one is to compete in a globalized world, one will have to pay attention to the fact that knowledge usually has a quite short half-life time. In other words it is necessary to see to it that all staff categories on the labour market get suitable, regular and high quality further education and training. Many of the newly industrialized countries have a lot of people with quite high formal education levels. However, without a suitable knowledge environment much of the acquired knowledge will quite fast be lost due to deficiencies in the infrastructure or due to many enterprises in the developing world having problems reaching critical knowledge level within many of their enterprises. Deficiencies in the physical environment may also quite often contribute to that fact. However, the situation just described in the developing countries in the previous sentences cannot go on for long. We simply have to continuously improve how different things are done. CVT plays an important role in this respect and it is the role of CVTS to describe that.

The costs and complexity of future surveys will increase if we collect information both from an enterprise at central level and from a sample of local units as well. We also have the problem of whether all participating countries have or will in the near future have access to a sampling frame that can be used for sampling local units. We can also not go on asking as many and to the enterprises difficult questions as we had in CVTS3. What will be important to do, is to include some questions about policies both in a questionnaire to enterprises at central level and at local unit level. When we ask for volumes and costs we should limit our questions to an absolute minimum. We should also concentrate on CVT-issues only and no longer ask for volume data by age groups. A survey directed to individuals quite often provides better information than what can be obtained in an enterprise survey about CVT-volumes by age, sex and similar qualitative background variables.

In a few years *ubiquitous* will probably be one of the most important buzzwords in discussions about the future and the technical development in general in different areas. With the word ubiquitous most people today mean a ubiquitous networking world. In short, one could describe this as a paradigm shift in technical development in connection to the ICT-development. The first wave in the ICT-development was characterized by mainframe computing (one computer shared by several individuals). The second wave or paradigm shift was characterized by the use of stationary and portable personal computers (usually only one computer per individual). Ubiquitous networking with the individual in the centre and surrounded by networking machines containing different kinds of computing capabilities can be said to characterize the third wave. In a ubiquitous society computing recourses are available for an individual in the form of a lot of different machines that usually are networking in some way. Such machines can for instance be personal computers, data terminals or personal computers connected to mainframe systems, digital cameras, pocket computers, copying machines, printing machines, mobile phones, different kinds of remote

controls, different production machines in the industry or offices etc. Networking between different machines is quite often wireless.

The new paradigm shift in the ICT-development described in the previous paragraph will of course not have taken over completely from the present paradigm, characterized by personal computers in different forms, before it will be necessary to perform a new CVTS. It will be of crucial importance to know and to be able to describe what is taking place within different enterprises even in this aspect. Already in connection to CVTS3 a couple of questions were co-ordinated with the Community Innovation Survey (CIS). In CVTS4 and in consideration to the development of the ubiquitous society I foresee that it will be even more important to co-ordinate both the contents and the wordings of some of the questions in CVTS and in CIS. It will probably be of vital importance to include one or two individuals with expert knowledge about CIS in the contract group and/or task force group in connection to the planning of the CVTS4.

In Sweden we have, with quite good results, since 1994 been able to connect a *cost model* to the Swedish Staff Training Survey. Our cost model is partly based on results from the Staff Training Survey, partly on results from surveys to different parts of the Swedish labour market about wages, salaries, other cost benefits and taxes and finally also partly on some assumptions. Extra questions in the Staff Training Survey only for use in our cost model are also included every five years or so in order to update some of the values used in the cost model. However, although we have a cost model for CVT-costs that function quite well in Sweden we cannot without reservations directly recommend a similar approach at the European level as the situation and the prevailing conditions in different member countries will probably differ quite a lot, something that will make it extremely difficult to develop such a common European cost model. Results from a common European cost model will probably not be either comparable between different countries or reliable enough if not quite a lot of recourses are put into such a project. In other words we cannot even recommend putting aside resources in trying to develop a common CVTS3 European cost model.

We simply cannot go on like before with the CVTS in the future. We will only get lower and lower response rates and the risk for biased results will increase all the time. It is therefore a clear risk that the relevance of the survey will be questioned if we do not make necessary changes to the survey. The irony, in CVTS3, that we quite often were met with in our contacts with the enterprises indicate that if we put more pressure on the respondents, for instance with a legal obligation for the enterprises to respond, we will risk getting fabricated results. A legal obligation for the enterprises is not a solution or even part of a solution to our problems as long as we still have all the difficult questions that we had in CVTS3. We simply have to better adapt the survey to future needs and to reduce the number of difficult questions. By doing that it will be possible to have a legal obligation to lean on and the enterprises will probably also feel more motivated to participate in the survey.

One important thing not to forget is the need to simplify regulations and reduce the burden on the enterprises. Questions about time spent and number of individuals involved in the process of answering our questionnaires like we had in our Swedish CVTS3 questionnaire should therefore be included in the common CVTS questionnaire in future surveys. However, when it comes to the CVTS in its present form one should not focus too much on results from similar questions. The most important thing should still be to try to get qualitatively more reliable information from the survey. In order to be able to do that it is imperative to try to simplify and put the right questions to the right level within the enterprises.

Important work concerning simplification of regulations and reductions of the response burden on enterprises are since several years performed by organizations like OECD, EU and UNICE and also within many individual countries. In connection to that work a cost model for the measurement of administrative costs has been developed by Netherlands. As mentioned earlier we cannot without reservations recommend development of a common European cost model for measuring CVT-costs. However, if one considers the importance of trying to simplify for enterprises it will perhaps even be necessary some time in the future to look seriously into this possibility.

There is a problem of being able to distinguish correct from wrong values especially when we have received quite low or no course costs at all from our respondents. This problem we foresee will increase in the future. An analogy, from the Netherlands, can perhaps best describe that. The Netherlands is geographically a small country with a high density population in many residential areas. Cable TV is therefore quite common in the Netherlands and telemarket actors are today strongly pressed by cable TV actors offering Internet, TV and telephony services in the same package offer. The cable TV actors often set low prices for their telephony services and count on making necessary profits on Internet and/or TV services. In other words, they count on making profit in other areas and can therefore dump or set low prices in areas where they come in as new actors. Similar problems exist probably in all union countries already today also within the area of education and training on the labour market. Equipment suppliers, for instance, quite often do not count on getting profits on their education and training activities but on other things like for instance the machines or other products on which management they provide courses. In general we can, in the future, expect that double, triple, quadruple etc. business transactions in different areas may lead to biased price levels also in the area of education and training on the labour market.

Extension of NACE and size coverage in CVTS - experiences in Bulgaria

Reni Dimtrova, National Statistical Institute of Bulgaria

I. The Bulgarian CVTS in enterprises prior CVTS3

The National Statistical Institute of Bulgaria (NSI) has been included in the European initiative of CVT observation in since 1999 (project CVTS2). By this time, the transition towards a market economy was still not completed. The system for vocational training had passed through many changes related to the restructuring of economy and the form of property. Some of the existing centres for vocational training in the country had been closed down, while others appeared in their places. That is why the government adopted several legal acts, regulating the activity within the vocational training. The participation in the project was a favourable possibility to supply the government and other concerned institutions with comparable and harmonized at European level statistical data on continuing vocational training in enterprises provided for their employees.

Meanwhile, because of too dynamic restructuring processes in the Bulgarian economy, interest and necessity of the government and all other actors on the labour market from information concerning Vocational Training in the enterprises increased. Taking into account these facts, the National Statistical Institute of Bulgaria carried out series of National CVTSs (in 2001, 2003 and 2005) in order to supply all involved bodies with fresh information.

CVTS 2001 – NACE enlargement

This survey was carried out together with the II quarter Employment Statistics Survey (Labour Costs Survey). A separate CVT questionnaire similar to the CVTS2 questionnaire was implemented. It covered a part of Employment Statistics Survey (ESS) population observed – enterprises with 10 and more employed from 24 NACE categories: sectors A, C – J, K+O, L, M, and N (about 13 230 enterprises), allocated in 92 strata. The scope of the survey was extended with 4 additional NACE categories: sectors A (Agriculture), L (Public administration, excluding defence), M (Education) and N (Health and Social Work). It included the reference year 2000 and the first six months of 2001.

CVTS 2003 – needs of enlargement

In 2003, a study of user satisfaction with CVTS2 results pointed out that the Ministry of Labour and Social Policy and other organizations took a major interest in defining the labour market policy as regards data on organization of the continuing vocational training in the micro-enterprises and some other fields of activity (Education, Health and Social Work, etc.).

Thus in 2003 NSI carried out another sample survey, following the CVTS2 methodology. The national survey was representative for the entire territory of the country and it covered enterprises (companies and organizations) of 24 NACE categories: A, C, D, E, F, G, H, I, J, K and O, L, M, N having a number of employees more than 5. The selected sample consisted of 4 000 units allocated in 96 strata. In comparison with CVTS2, the sample size was 33 % larger.

The **main goal** was:

- to provide current data about training policy of the observed units, fields of training, number of employees involved in training activities and others;

- to develop, to implement and to test some of the proposals Eurostat made for improving the next CVTS in compliance with the drafts of Commission Regulation, accordingly the contracted with Eurostat Grant clauses.

The CVTS 2003 was organized in the context of the LLL conception and included except the CVT which the enterprises had been organising for their employed in 2002 and Initial Vocational Training (**IVT)** measures as well, as an element of the development of the common vocational skills persons employed. The survey focus was on Continuing vocational training in spite of inclusion on IVT in the scope of the observation.

The new, that distinguished this survey from CVTS2, is as follow:

1. Enlargement the scope of the survey through including of four additional NACE categories: sectors A – Agriculture, M – Education, L – Public Administration and Compulsory Social Security (excluding defence) and N -Health and social work and the enterprises with 5 to 9 employed as well.
2. New questionnaire design

2.1. Incorporation a new section with questions, concerning IVT measures for persons who have a training contract (apprentices, trainees and others) ;

2.2. Reduced volume of the questionnaire by removing some non-obligatory CVTS2 questions and combine others, where it is possible

3. Different survey approach – obtaining information from other statistical survey (labour costs).

The questions about the number of employed, hours worked and labour costs due to its nature (necessity to make reference with other documents – accounting reports, etc.) required a longer time to be filled-in. By removing them from the questionnaire its size was reduced, as well as the time for filling-in. It contributed to the reduction the respondents' burden. Data about the total number of persons employed, total number of hours worked and labour costs of the employed for each enterprise found in the CVTS's sample are entered in addition from the file containing the annual data of employed, means for salaries and wages and other labour costs for the same enterprise.

CVTS 2005 - a transition from CVTS2 to CVTS3

The survey was carried out (reference year 2004) accordingly to the Grant conditions in the frame of MBP PHARE 2003. As in the previous (2003) national survey we have applied the CVTS2 methodological principles, but some recommendations of CVTS3 TF working group were took into account in the process of preparation.

The number of the sampled enterprises was 3 813 (27% more than CVTS2), distributed in 92 strata (23 NACE and 4 size groups).

In comparison with CVTS2 the scope of the survey in 2005 was extended in the following directions:

1. Extension of the survey subject to initial training

Inclusion of additional questions concerning Initial Vocational Training (IVT providers, participants in IVT, cost of IVT and others)

2. Extension of coverage of NACE sections L, M and N

By inclusion of these 3 sections into the survey scope we got the possibility to satisfy the users' needs and to supply the government with information about training policy in the schools, universities, public administration and public health services. On the other hand, including respondents in the survey, which are not business establishments, created some difficulties in relation to the form of the asked questions.

3. Extension of the coverage to enterprises with 5-9 employees

This on its side led to a significant increase of the sample size. Besides, the micro-enterprises are more mobile and to a certain degree it caused some difficulties to the interviewers.

4. Inclusion of additional variables

In order to increase the response rate by item, definitions of some questions have been clarified and complied with the national peculiarities at the most and the form and coding of some questions have been changed.

Some efforts were made in order to reduce response burden, to improve the data quality. In the process of the national CVTS preparation we succeed to implement and test the last methodological suggestions of Eurostat and the main topics of the Commission Draft Regulation. Additional questions related to the number of CVT courses in respect of forms of training and "Other forms of training" were included in the questionnaire. This way we obtained data, which gave the possibility to calculate some new indicators in comparison with CVTS2:

- total number of CVT courses;
- number of CVT courses per enterprise;
- average time spent in CVT course;
- average costs per CVT course.

In order to obtain more data on "Other forms" of training we have asked the enterprises to supply us with a list of these forms of training. The same has been asked about the rest of the qualitative indicators.

At the same time, in order to reduce the respondents' burden and to avoid the duplication of the statistical data gained from other surveys, the following data indicators for the enterprises were gotten from annual ESS for 2004: main sector of activity, number of the persons employed, paid working time (hours) and labour costs. Thus the questionnaire was shortened and the average time for filling in was reduced by 20%.

II. Assessment of the effects of the extension with regard to:

- **non-response rates**

The inclusion of additional NACE categories and micro-enterprises in the national surveys did not influenced negatively of the response rate by units – they were higher then CVTS2 and CVTS3 response rates. However the main part of the non filled-in questionnaires concerned the micro-enterprises. Non response rate was as follow:

CVTS2	– 10.8%
CVTS 2003	– 4.8%
CVTS 2005	– 3.5%
CVTS3	– 12.9%

- **quality and comparability of data**

The data comparability of national CVT surveys was provided by implementing same methodology (CVTS2).

The item non-response by some questions is not sufficient. For example when it concerned costs of training – micro enterprises and small enterprises often did not have a separate issue in their budget for training costs or about type of training costs.

- **costs aspects**

In fact the extension of the sampling frame in respect of NACE and especially with including the micro enterprises into the survey scope affected on the survey cost. It is very difficult to point the real part of the increased expenditure caused separately by size extension and by NACE extension. We estimated that the national survey costs increased approximately with the same percent as the sample size.

- **users satisfaction**

By conducting of several CVT surveys main users were provided with more completed fresh data about training policy in the micro enterprises and enterprises and organizations from the sectors C- K+ O and the Education, Health and Public Administration as well. The inclusion of some questions concerning IVT gave an overall picture on the training provision in the enterprises.

III. Needs and consequences of the extension of coverage with respect to NACE and size of the enterprises – recommendations for future CVTS4 survey

1. An overall assessment of the national CVTS results supports our position that there is a need to extend the scope of forthcoming CVTS4 **in relation to additional NACE** categories. The extension will ensure better satisfaction of the user needs in more detailed data on vocational training. In relation to the number of employed, by our opinion there is no need **enterprises with 5 – 9 employed** to be included in the survey scope. The obtained data did not differ significantly from same data about enterprises with 10 – 49 employed, and the micro enterprises have higher non-response rate.

2. After a critical assessment of the responses we have concluded, that it is not necessary to enlarge the list of “other forms” of training.

3. We suggest the **time lag** between the CVTSs to be 3 years.

4. We recommend that the size of the questionnaire to be shortened by removing the structural data.

Extension of NACE and size coverage in CVTS - experience in Bulgaria

The Bulgarian CVTS in enterprises prior CVTS3

Needs of the national CVTS

1. Dynamic restructuring processes in the Bulgarian economy:

- Development of the private sector – new training bodies have appeared
- Essential changes in the proportion of the economic activities: new activities, new products and technologies and new skills
- Implementation of the ICT – new professions
- European Integration of Bulgaria – commitments to European educational and training programmes and politics
- Changes in the national legislation concerning the VET

2. Information necessity:

- Government
- All bodies, participating on the labour market
- Educational and training organizations
- Trade unions
- Science institutions
- European Parliament end Commission
- Other international organizations

CVTS in Bulgaria following the CVTS2 methodology

- **CVTS 2001** – NACE enlargement

  This survey was carried out together with the second quarter Employment Statistics Survey (a separate CVT questionnaire was implemented). It covered the part of the ESS population observed – enterprises with 10 and more employed from 24 NACE categories: A, C – J, K+O, L, M, and N (about 13 230 enterprises), allocated in 96 strata. No deviations from CVTS2 questionnaire.

CVTS 2003 – Needs of enlargements

Necessity of topical information about:

- micro enterprises
- additional NACE
- IVT

1. Enlargement the survey scope in respect of NACE and size

CVTS 2003 was a separate sample survey (reference year 2002) implementing CVTS2 methodology and the recommendations of Eurostat's CVTS team.

The selected sample consisted 4 000 units (96 strata): enterprises with 5 and more employed and 4 additional NACE (A, L, M and N). In comparison with CVTS2 the sample size was 33% larger.

2. Extension of the survey subject to initial training

A **new section** with questions concerning IVT (enterprises provided IVT, participants, hours spent in training and cost of training) was incorporated in the questionnaire.

3. New methodological approach

In order to reduce the respondents' burden and to avoid the duplication of the statistical information gained from other surveys, the following data indicators for the enterprises were taken from annual Employment Statistics Survey for 2002:

- main sector of activity;
- number of the persons employed;
- paid working time (hours) and labour costs,

In this way, the questionnaire was shortened and the time, spent of filling in was reduced by 20%.

CVTS 2005 - a transition from CVTS2 to CVTS3

1. Enlargement the survey scope in respect of enterprises' NACE and size

A sample survey was carried out (reference year 2004) accordingly the Grant conditions in the frame of MBP PHARE 2003.

Sample size – 3 813 units (27% larger than CVTS2)

92 strata (23 NACE and 4 size groups).

2. Extension of the survey subject to initial training

- IVT activities
- Participants
- Hours spent in IVT
- Costs of IVT

3. New questionnaire design

- The questionnaire was shortened significant (removing the structural data)
- Inclusion a question related to the number of CVT courses in respect of forms of training
- Additional question about "Other forms of training"

Assessment the effects of the extension with regard to:

- **non-response rates**
  CVTS2 – 10.8%
  CVTS 2003 – 4.8%
  CVTS 2005 – 3.5%

- **quality and comparability of data**
  item non response – there are some questions with increased non response rate (for example costs of CVT in micro enterprises)
  the data of micro enterprises does not differ significant from same data about small enterprises

- **costs aspects**

– because of the extension the national CVTS's, it was about 25% more expensive than CVTS2.

- **users satisfaction**

– main CVT users were provided by more completed fresh information about training policy in the micro enterprises, enterprises and organizations from the sectors of the Education, Health and Public administration. The inclusion of some questions concerning IVT gave an overall picture on the training provision in the enterprises.

Needs and consequences of the extension of the coverage with respect to NACE and size of the enterprises – recommendations for future CVTS surveys

1. An overall assessment of the national CVTS results supports our position that it is needed to extend the scope of forthcoming CVTS4 **in relation to include additional NACE** categories. The extension will ensure better satisfaction of the user needs in more detailed vocational training information.
2. In relation to the number of employed, our opinion is that there is no need to include **enterprises with 5 – 10 employed** in the survey scope.
3. We suggest **the time lag** between the CVTSs to be set at 3 years.

CVTS3 - Pilot study report, UK

Anthony Clarke, Department for Children, Schools and Families, UK

Introduction

Countries participating in CVTS3 were required to sample enterprises employing 10 or more persons in the NACE categories C – K & O, within the main sample. However, as a pilot exercise countries were encouraged to extend coverage to include enterprises employing 5-9 employees, and/or NACE categories A, B, L, M, N & Q (NACE category P, domestic staff, being excluded on the grounds that it is of minor importance). Feedback from these pilot exercises will inform the feasibility of extending the coverage for CVTS4.

In the UK the main sample was extended as part of the pilot exercise to cover NACE categories A, B, L, M & N (Agriculture, hunting and forestry, Fishing, Public administration and defence, Education, and Health and social work). NACE category Q (Extra-territorial organisations) was excluded due to practicalities - they are not included in our business register, as well as their minor importance in relation to this survey. Results from these additional interviews are the basis of this report.

The UK sample was not extended to cover enterprises with less than 10 employees. Whilst there is much policy interest - and national surveys often cover smaller employers - it was felt more important to maintain continuity with previous CVTS studies and comparability internationally across countries, along with minimising cost and respondent burden.

Recommendations

1. Future **CVTS studies should be extended to include NACE categories A, B, L, M & N**. This being dependent on a sizeable majority of countries having a relevant sampling frame available, and noting Council regulation 2186/93 (see Annex B) whereby the inclusion of NACE categories A, B & L within business registers is optional. In particular, categories L, M & N are of great policy interest due to the proportion of enterprises and employees that they cover.

2. **NACE categories P & Q should continue to be excluded** from the survey due to their minor importance in relation to CVT as well as difficulties in producing a reliable sampling frame.

UK results

A total of 895 interviews were conducted in the pilot sample, from a population of 35,500 enterprises. The importance of these NACE categories is demonstrated by them covering 5 million employees in the UK, equal to almost a quarter (24%) of the total employment for enterprises with 10 or more employees. Overall, the pilot sample represents 17% of enterprises within the UK - mainly due to NACE category N: have health and social work, which represents 12%.

Number of enterprises in the population (IDBR[1])

NACE category	10- 49	50-249	250+	Total	% of CVTS3 population
A: Agriculture, hunting and forestry	3,032	250	26	3,308	2%
B: Fishing	63	5	2	70	0%
L: Public administration and defence	180	134	429	743	0%
M: Education	3,886	1,570	855	6,311	3%
N: Health and social work	20,895	3,109	1,024	25,028	12%
TOTAL - PILOT	**28,056**	**5,068**	**2,336**	**35,460**	**17%**
TOTAL – CVTS3	174,149	31,860	8,416	214,425	100%

1 - The sample was drawn from the Inter Departmental Business Register (IDBR), details of which are in Annex A.

The pilot sample accounted for a fifth (21%) of interviews in the UK. Education and Health and social work both accounting for 8% of interviews.

Number of enterprises in the sample (CVTS3)

NACE category	10- 49	50-249	250+	Total	% of CVTS3 population
A: Agriculture, hunting and forestry	75	19	5	99	2%
B: Fishing	6	0	0	6	0%
L: Public administration and defence	24	21	68	113	3%
M: Education	139	91	90	320	8%
N: Health and social work	175	121	56	352	8%
TOTAL - PILOT	**419**	**252**	**219**	**890**	**21%**
TOTAL – CVTS3	1918	1507	835	4,260	100%

We were not able to calculate response rates high individual NACE categories, and so for an indirect measure of response we instead look at the coverage rate as calculated by dividing the number of achieved interviews by the total number of enterprises in

the population. The overall coverage rate of 3% for the pilot was higher than the overall total of 2% for CVTS3. However, this varies widely between NACE categories - from 1% in N: Health and social work to 16% in L: Public administration and defence.

Coverage rate - proportion of all existing enterprises interviewed

NACE category	10- 49	50-249	250+	Total
A: Agriculture, hunting and forestry	3%	4%	13%	3%
B: Fishing	11%			11%
L: Public administration and defence	16%	16%	15%	16%
M: Education	4%	5%	10%	5%
N: Health and social work	1%	2%	4%	1%
TOTAL - PILOT	**2%**	**3%**	**8%**	**3%**
TOTAL – CVTS3	1%	3%	6%	2%

The pilot sample represents 24% of employment, mainly due to NACE category N: Health and social work (12%), along with significant contributions from categories L: Public administration and defence, and M: Education, both 6%.

Number of employees represented by each industry sector (CVTS3)

NACE category	10- 49	50-249	250+	Total	% of CVTS3 population
A: Agriculture, hunting and forestry	60,000	47,000	22,000	129,000	1%
B: Fishing	1,000			1,000	0%
L: Public administration and defence	3,000	17,000	1,137,000	1,157,000	6%
M: Education	75,000	204,000	1,013,000	1,291,000	6%
N: Health and social work	361,000	605,000	1,410,000	2,377,000	12%
TOTAL - PILOT	**500,000**	**873,000**	**3,582,000**	**4,955,000**	**24%**
TOTAL – CVTS3	3,469,000	4,230,000	12,910,000	20,609,000	100%

Analysis by NACE category

A: Agriculture, hunting and forestry

NACE category A accounts for only 1% of the population of enterprises in the UK, as

well as 1% of employees. Nationally, it is common practice to combine results from NACE category A with NACE category B (Fishing).
The coverage rate (ratio of interviewed enterprises to all enterprises in the population) of 3% was slightly above the overall average for CVTS3 of 2%. This suggests there weren't particular problems gaining responses from this industry sector.

B: Fishing
NACE category B accounts for a very small population of enterprises – less than 1%. With only five enterprises of size 50-249 in NACE category B and only two enterprises of size 250+ this presents problems of disclosure. To solve this problem, results from NACE category B should be combined with NACE category A. With a coverage rate of 11% response rates were good.

L: Public administration and defence

Although only a small number of enterprises (1%) are in this NACE sector, they include some very large employers and so account for a disproportionately large percentage of overall employment - 6%. Employers in this sector had the highest likelihood of providing training, although this will be related to their relatively large size. The coverage rate was very high at 16%.

M: Education

The Education sector accounts for 3% of enterprises but because some of these are large it covers 6% of employees. These employers were more likely to train than the average across the whole CVTS sample, although again this will be related to a relatively large size. The coverage rate of 5% was above average.

N: Health and social work

The Health and social work sector covers a very large proportion of UK enterprises - 12%, also accounting for 12% of employees. Provision of training courses was again above average. The coverage rate of 1% was below average, with the rates for small, medium and large enterprises all being below average.

Annex A: Inter Departmental Business Register (IDBR)

The IDBR is a list of UK businesses maintained by our National Statistical Institute (Office for National Statistics) and combines the VAT based business register and the employment statistics system. It complies with European Union regulation 2186/93 on harmonisation of business registers for statistical purposes.

What businesses are included?

It covers all parts of the economy, but misses some very small businesses (self employed and those without employees and low turnover) and some non-profit making organisations. There are around 4.4 million businesses in the UK of which 2.1 million are on the IDBR. The IDBR provides nearly 99% coverage of economic activity.

How is it updated?

Using data from 3 main sources:

- **Value Added Tax (VAT)**
  Detail of businesses registered for VAT.
  Covers 1.7 million traders.
  Updated daily.
- **Pay As You Earn (PAYE)**
  Details of employers with employees in PAYE schemes.
  Covers 1.1 million employers.
  Updated quarterly.
- **NS Surveys based on the IDBR**
  1.4 million survey forms sent to 273,000 businesses in 2005
- **Other sources include**
  Dun and Bradstreet - commercial database of businesses
  Companies House - the official UK government register of UK companies

What type of business units does it hold?

- **Administrative units**

  VAT traders and PAYE employers information forms the basis of the IDBR
- **Statistical units**

  A group of legal units under common ownership is called an enterprise group. An enterprise is the smallest combination of legal units (generally based on VAT or PAYE) which has a certain degree of autonomy within an enterprise group. An individual site (factory, shop etc.) in an enterprise is called a local unit.
- **Reporting unit**

  Reporting units hold the mailing address to which the survey forms are sent. The form can cover the enterprise as a whole, or parts of the enterprise identified by lists of local units.

Annex B: Council Regulation (EEC) No 2186/93 of 22 July 1993 on Community coordination in drawing up business registers for statistical purposes

The inclusion of:

- enterprises the main activity of which falls within Section A, B or L of NACE Rev. 1,
- legal units responsible for them,
- local units dependent on them,

shall be optional.

Continuing Vocational Training Survey Workshop

How to collect and process CVTS data more efficiently - practices in Germany

Daniel Schmidt, 30 June 2009, Cedefop, Thessaloniki, Greece

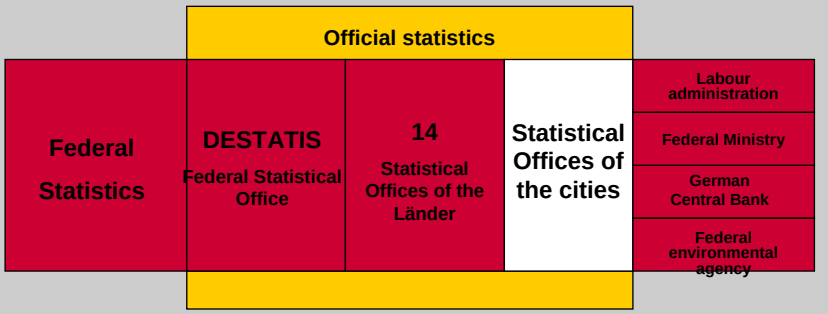
Outline

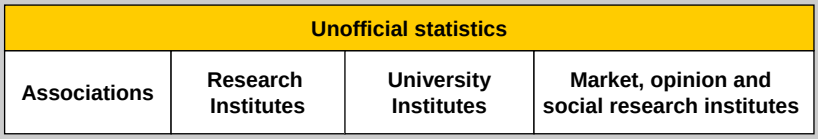
1. **The German Federal Statistical System**
2. **CVTS3 in Germany**
3. **Possibilities for collecting CVTS data in a postal survey**
4. **Data editing and control**
5. **Efficient data processing**
6. **Preview of CVTS4 in Germany**

Federal Statistical Office

DESTATIS wissen.nutzen.

Functional organisation of the Federal Statistical System

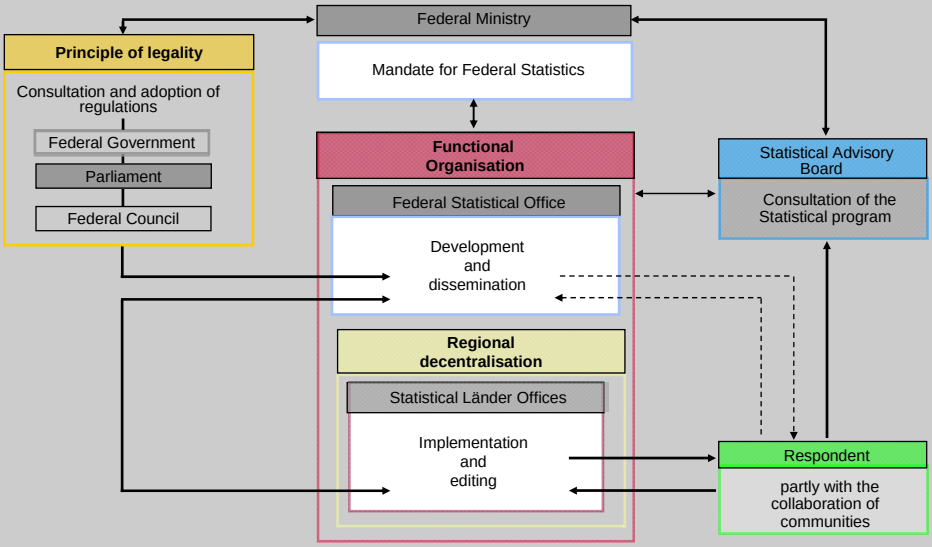




Federal Statistical Office

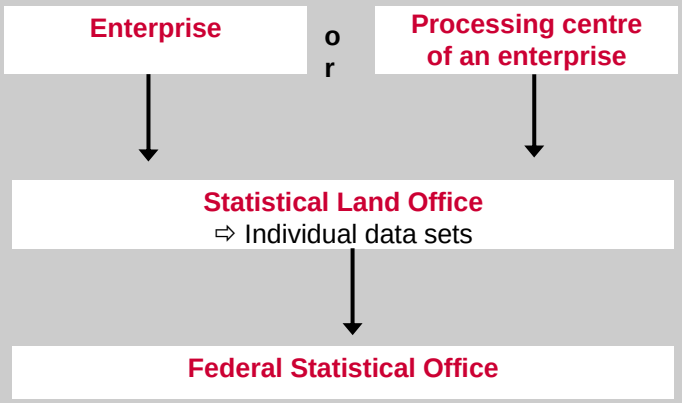
DESTATIS wissen.nutzen.

Course of Federal Statistics



CVTS3 in the previous framework

- **Statistical data flow**



CVTS3 in Germany

- CVTS3 was carried out in 2006
- The national legal basis was Commission Regulation (EC) No 198/2006 of 6 February 2006 implementing Council Regulation (EC) - No 1552/2005 - on statistics relating to vocational training in enterprises associated with the German Statistical Law
- It was a sample survey covering enterprises of NACE categories 10 - 37, 40, 41, 45, 50 - 52, 55, 60 - 67, 70 - 74, 90 – 93 with 10 or more employees (sample size 10 615 enterprises)
- The survey was non-compulsory
- The data collection was based on a “pure” postal questionnaire
- Due to a tight European time schedule, only a one stage data collection could be realised

Preparation of the fieldwork in a postal survey

- Development of a list with addresses of the sampled enterprises and allocation of a clear ID-number (important in reference to identify an enterprise in all phases of the survey)
- Information of enterprise associations in the different economic sectors with the primary goal to reach a good climate for the survey
  - In some cases the contacted associations informed directly their members
  - In others they published an article in their specific journal
- Information of the sampled enterprises about two to four weeks prior to the sending date of the questionnaire
- Sending the questionnaires with a deadline, that means enterprises should have responded by this deadline

Importance of reminders in a postal survey

First postal reminder:

- First postal reminder to all enterprises that did not answer after the deadline specified in the initial shipment
- Only a letter with a new deadline without a new questionnaire

Second postal reminder:

- Second postal reminder to all enterprises that did not answer to the first reminder after the deadline specified in the first postal reminder
- Letter with a last deadline and once more the questionnaire

Reminder by telephone contacts:

- Finally we got back to a certain number of enterprises by telephone contacts
- Here it was important to contact primarily enterprises in stratums with a very low response rate and huge enterprises with many employees, and moreover, those enterprises which had promised to sent the questionnaire and had requested a later date for sending it back

Response organisation with a list of response

List of response

Number of questionnaires: ...
Date: ...

	ID number	date of receipt	Unit response	Overcoverage							Refusals				notes
				not in sample frame	new economic focus	Liquidation or insolvency	Split or merged enter-prises	left Germany	double shipped	other reasons	real refusals	unknown, adress unknown	unactive	other reasons	
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1															
2															
3															
4															
5															
6															
7															
8															
9															
10															
11															
12															
...															

Data editing and control

- Editing and control of incoming data includes validity checks for the variables and consistency checks between different variables
- In most cases, call backs are the only way to find out missing data or to correct values

The German approach:

- Directly after receiving the completed questionnaires an intensive manual checking of their contents is performed
- Development of a detailed handbook for the staff that was involved in the editing process
- Contents of the handbook:
  - How to find out missing data and inconsistencies
  - How to correct them
  - How to code the answers

Relevance of the national editing handbook (1)

- The handbook contains general instructions and especially comprehensive explanations for each variable.

General instructions:

- Identified missing values or implausible answers should be clarified with the respondent by telephone contact
- This telephone contact should be done only after manual checking of the whole questionnaire
- If an enterprise could not provide figures for variables with reference value, also shares should be accepted (e.g. paid working time spent on CVT courses by fields of training or costs on CVT courses)
- If the additional contact did not lead to clarification the variable had to be coded with “No Answer”
- Data like “-” nothing had to be recoded into zero “0”

Relevance of the national editing handbook (2)

Example showing the contents of the editing handbook: explanations for the variable persons employed by the enterprise by age groups.

1. The distribution of persons employed by the enterprise on 31.12.2005 is absolutely necessary.
2. For one of the three age group variables a value must be available. A value for the other variables is not absolutely required. The overall sum must be equal to persons employed by the enterprise on 31.12.2005 . In this case all variables without a value are to be coded with zero “O”.
3. Enterprises must be called back, if they did not distribute the persons employed by the enterprise on 31.12.2005.
4. If an enterprise could not provide values for at least one variable, the code 999999 “no answer” must be entered for all three variables.
5. If an age group is not available in the enterprise, it has to be coded with zero “O”.

Data processing

- For automatic checking, we used a BLAISE application which was mainly based on the checks included in the CVTS3 checking rules with some national additions
- The user interface of the application is adapted to the postal questionnaire
- The figures provided by the enterprise are entered via the user interface into the application
- If a figure conflicts with the checking rules a mandatory instruction or a warning is issued

Problems of non-response in a non-compulsory survey

- Reduction of the number of sampled enterprises by non-response might cause less precise estimates for key variables regarding the different strata
- Non-response can lead to over- or under representation of some groups in the population

  $\Rightarrow$ introduction of bias

Possible Solutions:

- Re-weighting is reducing the unit non-response
- Imputation is reducing the item non-response
  - Missing values should be imputed, if all kinds of recommended efforts have been made to receive missing values from the enterprise

Methods of imputing missing quantitative variables

How to impute missing values for the quantitative variable total hours spent in CVT courses, for example?

Data sets required for imputation:

1. All data sets in a stratum with participants on CVT courses (missing value for total hours spent in CVT courses)
2. All data sets in a stratum with participants on CVT courses and total hours spent in CVT courses

Method of imputation:

To impute the total hours spent in CVT courses calculate the hours per participant from (2) and multiply by total participants for each enterprise in (1)

Imputation of qualitative variables

To reduce the item non-response of qualitative variables we used the method “Sequential Hot Deck procedure”, that means:

- Development of a specific national MS Excel-macro for these imputation procedure
- Preparation of fixed reference classes (NACE20 and six size classes)
- In each reference class the existing values of a qualitative variable were assigned sequentially to missing values
- If the first data record is missing it would be substituted with the initial value
- Arrangement of data records takes place in a random way
- Imputed values would be checked for plausibility

Preview of CVTS4 in Germany

- CVST4 will be carried out in 2011
- Enterprises will have different options to provide their data
  - a postal questionnaire
  - modern online reporting methods like internet-based data collection which allows form-based input of data and uploading of files
- Whether the survey can be made for the enterprises compulsory will be discussed with the Federal Ministry of Education and Research
- It is currently envisaged to extend the coverage to the 3 economic sectors public administration, education and health
- The implementation of a two stage data collection is not envisaged

Thank you for your attention

Daniel Schmidt
Federal Statistical Office, Wiesbaden, Germany
Division VI B - Education, Research and Development, Culture
Tel. +49 (0) 611 / 75 - 43 75
Fax +49 (0) 611 / 75 - 39 97
E-mail cvts@destatis.de
daniel.schmidt@destatis.de

CVTS3 experiences in Cyprus and ways to improve the survey

Aristi Ioannidou, Statistical Service of Cyprus

The continuous vocational training survey was conducted in Cyprus for the first time in 2006 with reference year 2005. Its main purpose was to collect data on vocational training provided by the enterprises for their employees. The implementation of the survey was based on Regulation (EC) No 1552/2005 of the European Parliament and the Council which sets the legal frame for carrying out a Continuous Vocational Training Survey in all member states and producing harmonized data on these issues.

Obstacles encountered and overcome

As this was the first Continuous Vocational Training Survey in Cyprus a number of problems and difficulties were encountered. Some of these were the following:

a. Response

The first issue that we had to deal was the response. Of the total sample of 1.322 enterprises, 1.016 questionnaires were finally completed, which represent 77 % of the sample, and 27 % of all enterprises with ten or more persons employed in the Cypriot economy. Among the 306 questionnaires which were not filled, 192 were due to refusal by the enterprise or had severe missing data (i.e. with at least half of the variables not completed), while 79 questionnaires referred to enterprises with less than 10 persons employed, 13 referred to enterprises which had closed down, 6 referred to enterprises which did not work for the full year of 2005, and 6 enterprises were not located.

Additionally, due to the fact that Cyprus, is a small country with a relatively small number of enterprises, the burden of supply of statistical data on enterprises and particularly of large enterprises (employing ten or more employees) which participated in the CVTS, is too high. Several enterprises complained that the Statistical Service is putting a lot of burden on them as they participate in more than one survey at the same time. This situation led to a higher non-response rate than expected**.**

b. Incomplete records

Unfortunately, there were cases where the disorganized information systems did not allow the respondents to give us the details we needed for filling in the questionnaires properly, i.e. many enterprises did not have good and detailed records about training courses that might have been offered. For the same reason there were cases where enterprises, gave us information that proved not to be fully correct. For some variables which were not core variables, imputation of missing data was possible but unfortunately this was not always the case.

When there was not a very high rate of missing data, estimations were performed using rates from the other completed questionnaires of similar characteristics, i.e. such as economic activity, size and region.

c. Number of employees

Sometimes, the number of employees on the 31st of December is quite different from the number of employees who work during the whole year, especially in the hotels and restaurants industry, where a lot of employees work during the peak season, and fewer employees during the off-peak season. Consequently, enterprises in this industry were often employing less than ten persons on 31st of December, so if the survey took only into account the number of employees at the end of the year, as requested by Eurostat, then not only some

enterprises would have been under-represented, but also some industries too. In order to overcome this problem, CYSTAT decided to consider the average number of employees for the full reference year as the size of the enterprise and not just the number of persons employed at the end of the reference year. The main reason for doing so is that if the number of employees at the end of the year was used, this would be incompatible with the total labor costs of all persons employed during the reference year which is reported in question A5. We believe that in the next CVTS the average number of persons employed for the whole year should be adopted.

d. Group of companies

Another issue was the case of enterprises belonging to a group of companies, which usually means that the continuous vocational training is a part of the whole group, and not the enterprise itself. So it was difficult to isolate the data for the company that was selected in the sample. In such cases result was made to estimation, usually proportion to the total number of persons employed.

e. Participation in CVT courses

Furthermore, measurement errors appeared when we tried to fill in Section C of the questionnaire. In big enterprises, the number of employees who participated in CVT courses was sometimes quite large, and it was then particularly difficult to know whether an employee had participated in more than one CVT course or not. Also it was difficult to separate the total hours spent on all CVT courses into the different types of training and between men and women. Additionally, enterprises encountered difficulties with breaking down the total hours spent in CVT courses by working time and non-working time. As a result in these cases, where completing questions was time consuming, respondents gave incomplete answers.

To overcome the above problems to a significant extent, what we need to do before the implementation of the next survey is to inform in advance the enterprises with a letter on the kind of data that will be needed in order to complete the questionnaire. This letter needs to be sent by the end of 2009 and before 2010 so the enterprises have in mind what information we will seek in 2011 when the survey will take place and we will visit them for the interview. In this way, they will be more prepared about the needs of the survey and they will hopefully keep good records about the training policies of their enterprises in a more detailed form.

We believe that this action will reduce response burden on one hand and increase the response rates on the other hand. When the data is easily accessible, enterprises are more willing to fill in or answer a questionnaire. Additionally, this will reduce the cost of the survey for the Statistical Service since the questionnaires will be filled in easier and quicker by the interviewers. Moreover, when the questionnaires are more complete, checkers will need less time to edit them and will not have to re-contact the enterprises to ask for clarifications or fill in any missing data or at least give their estimations. In addition, when less time is needed for checking, less time will be needed for preparing and disseminating the results on time.

How can data quality be improved?

Learning from French experiences for CVTS-3

Agnès Checcaglini, Céreq, Centre for Research on Education, Training and Employment, France

Preliminary remarks:

This paper uses the French work carried out for a contribution in the "Colloque francophone sur les sondages" in November 2007, which has been published : CHECCAGLINI A., RECOTILLET I. et ROUSSET P. (2008), « Garantir la qualité de données d'entreprises françaises pour une enquête européenne » in GUILBERT Ph., HAZIZA D., RUIZ-GAZEN A. et TILLE Y. (dir). Méthodes de sondage : applications aux enquêtes longitudinales, à la santé, aux enquêtes électorales et aux enquêtes dans les pays en développement, *Dunod, 380p. Some conclusions have also been formulated in the French quality report for CVTS-3 and are from the Céreq team in charge of CVTS-3.*

1. Difficulties of collecting enterprise data

The French data collection was organised around two ways, depending on the type of questions. Answers to sections in which we find quantitative questions (sections A, B, C, F) were possible by phone or web in order to give respondents more flexibility. All qualitative questions (section D or E of the questionnaire) were asked by phone so that all enterprises had in contact with an interviewer at least once during the period of the survey. We tried think of the best way to reduce the burden for the French enterprises because, besides the usual difficulties of lack of enterprise availability for phone surveys, the collected data were numerous and responded to European definitions which required adaptation of the data usually collected in France from enterprises. CVTS-3 was introduced to the CNIS (French National Council for Statistical Information) and the Comité du Label (Label Committee), which is made up of institutions and social partners, with its protocol and the mixed method to collect data by web and by phone. The Comité du Label awarded the survey the 'General Interest and Statistical Quality' label and made it compulsory. This most probably contributed to obtaining a good global response rate.

1.1. Assessment of the web use

A web site was made for the data collection of CVTS-3 in France. The website introduced the survey and the Céreq institution. Enterprises could find official documents, the questionnaire and there was a personal access to field the questionnaire. The possibility of using the web was very well received by the enterprises when the interviewer offered them this opportunity by phone. The website was seen by the enterprises as a real effort to reduce their burden taking into account the fact that enterprises are confronted with quite a high number of statistical data collection surveys and to give them the opportunity to fill in the questionnaire at a convenient moment.

But the web site use was not as important as expected. In total the enterprises that connected to the website represented 17% of those who received codes to connect. Only one enterprise out of two connected filled in the questionnaire on the web site and completed the

interrogation by phone as planned. Some others answered only by phone, some partly answered by web but didn't agree to give more time by phone to complete the questionnaire, and some after the web site visit didn't agree to answer. The enterprises which used the web site didn't differ from the others for a particular size or NACE.

Through this experiment, it seems that although enterprises come out in favour of the introduction of new technologies, in practice they rarely use them, and more for inquiring or for preparing the interrogation than for recording their answers. From a technical point of view, a double data collection method was quite complex and risky in terms of data management and of sequence of different stages of enterprises contacts. This complexity wasn't compensated for by a satisfactory user rate.

1.2 Extension of the data collection period

The calendar for the data collection period was chosen taking into account the closing date for the enterprises continuing vocational training accounts and the summer holiday period, unfavourable for enterprise interrogation. But among others reasons because of the complexity of two collection methods, at the end of the initial collection period (14th July 2006), only around 3.500 enterprises had responded. This number was lower than expected and it was decided to extend the data collection period by one month with some methodological adjustments.

With the additional month (September), the number of respondents increased noticeably and reached almost 4,800 enterprises. The sub-contractor was requested to make some protocol changes. In this second round, enterprises could answer by phone only, although they were allowed to use the internet website to download the quantitative parts of the questionnaire and a target letter that would help them to fill in the questionnaire. In addition, in order to improve the response rate, the month of August was spent looking for phone numbers of enterprises for which we had an incorrect number, a serious problem in this kind of survey. Moreover, many enterprises agreed to participate in CVTS-3 during the first contact but afterwards were difficult to reach. They constituted a group of potential respondents. Less than 15% of the initial sampling remained difficult to reach (hidden refusals, postponed appointments or unanswered calls). Overall, 78% of respondents agreed to participate on first contact and 23% of those with whom the first phone contact wasn't conclusive later responded. The pre-contact stage bore fruit but wasn't enough. During the additional month, the decision was made to persist with attempts to contact enterprises who were difficult to reach during the first stage. In this way, the total response rate was increased and the non-response bias limited.

1.3 The proximity of administrative data: an aid for the survey

In France, enterprises must file tax returns regarding the funding of continuing vocational education for their employees and that constitutes the national administrative and fiscal source (24.83). In previous CVTS in France, quantitative data were extracted from the administrative and fiscal source and were completed with a phone interview to be in conformity with the European questionnaire. For CVTS-3, the enterprises were interviewed for all the variables. But to prepare the data and to answer the survey, they used the same documents as those they usually use to fill in their tax returns. Moreover both calendars of data collection were coordinated to reduce enterprise statistical burden. But, sometimes, this proximity could also be a trap because the definitions were not exactly the same and the fiscal figures had to be adjusted to the European definitions.

More generally the existence of this database was very helpful in assessing the global data coherence and quality.

2. Efforts to ensure data quality

Different initiatives were undertaken to improve the enterprise data quality with CVTS-3. Cereq set up measures in three ways: before the data collection, during the live interviews and during the whole period of the fieldwork with the control of collected data. The nature of those measures was from two types: a qualitative accompaniment and automatic control procedures. To ensure the optimum organisation of the CVTS-3 in France, a steering committee was set up which met on three occasions. The committee was composed of institutional and academic representatives, enterprises, and social partners. It aimed to convince the social partners of the usefulness of CVTS-3, and to anticipate any difficulties for French enterprises in understanding or interpretation of the standard European questions. It also reflected on the best way to formulate the questions and of looking for the administrative data necessary for the CVTS, and decided on the French questions which needed to be added.

2.1. The qualitative accompaniment

Qualitative procedures were undertaken to ensure data quality during and prior to the interrogation. One of the most important difficulties was identifying the right person to answer the survey. It meant the person who was able to speak about CVT policy and CVT budget. After which, we indicated that he/she had to prepare the interview. He / she became our “survey correspondent” who was also in charge of collecting the information from his/her enterprise like labour costs or the yearly number of hours worked or variables about initial vocational training. The aim of this step was to minimize incorrect answers or non responses.

Otherwise, during all the survey period, enterprises could connect to the survey website where they could find official documents, European definitions of the variables, the quantitative part of the questionnaire. A toll-free number was available for respondents and was located in the subcontractor's call centre. A person was assigned to answer calls, but for very difficult questions the call was transferred to the Cereq staff member present at the call centre. The possibility of using the toll-free number to complete their answers was also offered to the enterprises, and the call was then transferred to the interviewers. The toll-free number was used 37 times to complete answers, and 95 times to request clarifications about the questions. In total 2,206 enterprises used this number between May and October. The majority of calls concerned requests for more information about the survey and dates for call appointments.

For the interviewers, two accompaniments were organized: regular specific training and technical pieces of information on their computer screen to ensure the data accuracy and a harmony in the pieces of information given to the enterprises. So Cereq regularly organized training courses for interviewers and a permanent staff member from Cereq worked with the subcontractor during the entire data collection period. She was especially trained for the needs of the survey and was able to answer the majority of questions raised, in particular those regarding the differences between French and European definitions. She listened randomly to calls and organized, where necessary, short individual or collective briefings. She was also able to speak directly to the enterprises, to explain European definitions for example.

2.2. Automatic monitoring procedures

CATI and CAWI technologies were used with a lot of checking rules among other Eurostat checking rules. Some constraints were also introduced. For example, for the core variables missing data was not possible. The enterprise had to give an answer otherwise the interview was stopped. For the key variables, the interviewers had instructions to insist. They also offer the enterprises the possibility of using the toll-free number to complete their answers. With the automatic filtering and immediate data checking, the data capture errors and global incoherence was minimized by these technologies. For a lot of variables it was real time processing where the data were tested and verified. If some data weren't acceptable for the test, the interviewer could insist to be sure that there was no mistake and the questions had been really understood by the enterprise. On each screen we had also the possibility of adding some definitions and details to help the respondent. Each checking rule activated during the interview was recorded on the final file. This offered the possibility of evaluating the global quality of each questionnaire filled in.

A frequency analyzes of the check activation showed three types of difficulties. First with 23% of control activation, a divergence was noticed between the number of people in the enterprise in the sampling source and the number given by phone. Second, some incoherence has been detected in the variables linked together. For example if the number of hours worked was impossible with the number of people declared. Third, the inventory difficulties for some information were revealed when differences were noticed between the sum of hours per provider and the total number of hours.

So quality checks were specified at macro and micro level. At macro level, the data was compared with other data sources by enterprise size. The most important was the fiscal source (employers' tax declaration n°24-83) about CVT. Even when the field was not exactly the same one, the averages and the statistical distribution seemed to corroborate for the total labour costs, the rate of enterprises offering CVT courses for instance. The average of labour cost and total of hours worked per activity sector were compared with national data.

Conclusions for future CVTS:

With these different experiments (preliminary contacts with the enterprises, the possibility of answering partly on a website, the use of CATI and CAWI technologies, the free call telephone number, the training of the interviewers…) we learnt a lot of things for the future survey. In conclusion, we can list the different aspects which will be conserved for the next CVTS or probably won't be.

Some results are negative or didn't have enough impact on the response rates. Some measures undertaken have to be develop such as:

- before collecting data, the preliminary contacts with the enterprises : it seems to be very important to encourage and convince the enterprises to better prepare the interview
- the possibility of answering partly on a website wasn't satisfactory from the lack of use of the web site viewpoint
- the coordination and the sequence of the different collecting steps to reduce the data collection's period

Among the positive results, we are satisfied with:

- the compulsory label of the survey in France
- the data collection by phone
- the continual quality control of collecting data with the use of CATI technologies. The data set included the control variables which indicate if checking rules had been violated.
- the preliminary contacts with the enterprises
- the persistency with attempts to contact enterprises who were difficult to reach during the first stage. In this way, the total response rate was increased and the non-response bias limited.
- a free call number for the enterprises about the survey during all the collection period
- the interviewers training and monitoring
- the CVTS data collection at the same period of the year as the fiscal data about continual vocational training (yearly employers' tax declaration n°2483) for which the French enterprises are accustomed to preparing all training documents.

CVTS 3 experience in Latvia and ways forward to improve the data quality

Anita Švarckopfa, Central Statistical Bureau of Latvia

In 2006 the Central Statistical Bureau of Latvia carried out the second survey "Continuing Vocational Training Survey" (CVTS3), in which information was compiled on vocational training of employees in 2005. The sampling size comprised 3 935 enterprises. During the data collection stage, all attempts were made to reach a high response rate. Particular attention was paid to acquisition of data from the large and medium-size enterprises. The final response rate was equal to 94.1 %. The response rate was 94.4 % for large enterprises and 96.8 % for the medium-size enterprises. Unit non-response rate was low. To carry out the survey, the whole methodology developed by the Eurostat was used. Information acquired shows that only 36 % of enterprises participate in any kind of vocational training. Thus, it may be concluded that, for further surveys it would be necessary to improve data collection at the enterprise level and contacts with respondents and information quality.

Survey quality improvement directions:

1) Additional attention must be paid to the motivation of the enterprises and respondents. Motivation of respondents to compile information on training of employees in enterprise by informing them about the legal basis of the survey, the importance of the survey policy in general and its possible importance at national level. Representatives of the companies should be involved in the development process of the survey conception, definitions and questionnaire forms of the survey. At the time when field work is started, it would be useful if respondents received brief informative leaflets containing information on previous surveys.

During the survey, it was concluded that enterprises have not collected and compiled all information necessary for the survey, and that, in turn, burdened respondents to fill in the questionnaire forms. Taking into account the fact that some enterprises did not keep accounts on the number of employees who had taken part in vocational training, it was not feasible to break down the employees by the form of training and age group; there were no accounts about the hours spent on training during paid working time, etc. Not all questions were filled in. Due to the afore-said reasons, 20 enterprises gave no answers to items B2a to B2e, C1, C2 C3, C4, C5, C6 and C7.

Organisation and funding of vocational training, as well as administration of information collection and compilation differs in various enterprises depending on their size. Only 2 % of small companies have a training budget, but among the large companies, 28 % have a training budget. In large enterprises, information on vocational training of employees is more formalised and better organised than in small companies.

Data quality was influenced by the turnover of employees in enterprises. Respondents have difficulties to remember/find information on the year 2004, items B3a and B4a. The main reason for it mostly was the rapid turnover of employees in enterprises.

2) Methods of data collection. Enterprises could submit the filled questionnaires by sending them by mail, bringing them in person, or filling them electronically and submitting via Internet. When we compared the questionnaires brought in person with the questionnaires received by mail, we came to the conclusion that the quality of the first was better. The questionnaires collected face-to-face were of the best quality. Very seldom mistakes e were found; the most typical of them – items B1 and B2 were filled incorrectly, respondents mixed CVT courses with other forms of CVT. In order to obtain information of higher quality and not to increase survey costs, it would be useful to use a mixed mode of data collection, postal approach for smaller enterprises and face-to-face approach for larger enterprises.

In order to ensure the quality of the information compiled during the conduction of the survey, CSB staff members/data collectors, who in their every day work have contacts with enterprises of the corresponding branch, were involved.
To handle respondents' and interviewers' questions and problems a special help-desk was formed.

Additional instructions for data collectors were prepared on how to check the questionnaire. The data collectors were taught how to contact, motivate and encourage enterprises to give answers to the questions of the questionnaire.

The process was co-ordinated and administered by the Central Statistical Bureau. To be able to improve the methodology of such kind of surveys in future, the data collectors were instructed on how to register the respondents' questions and objections. Due to a heavy burden, data collectors rarely performed this documentation, thus, information on respondents' problems related to filling in of the questionnaire form is only partial. To have enough time to contact the respondents, the planning of data collectors work for the preceding period of CVTS 4 data collection should be improved.

3) Respondents in enterprises. During the survey, it is essential that questionnaire form in enterprise is delivered to a person informed about the vocational training of employees. Along with the questionnaire form, enterprises receive an informative letter containing information on the aims of the survey and a short description of the questionnaire form. The questionnaire form contains also information about the respondent: name, telephone number, and e-mail address. To acquire as precise information as possible on the person filling in the questionnaire form, it would be necessary to know the occupational position of the person filling the form, and this information may also be included in the CVTS 4 form.

4) Response burden and response rates and quality. To reduce the response burden, the following information was pre-printed in the questionnaire: Principal economic activity of the enterprise, Number of persons employed, Number of hours worked in the year 2005, Total labour costs of persons employed. The CSB is of the opinion that in such a

way the burden of enterprises is reduced and data comparability is ensured with the data included in other surveys. It should be mentioned that it was possible to pre-print this information in 2 725 or 69 % of questionnaires out of the total number of enterprises included in the sample (3 935 enterprises). Other enterprises had not been included in other CSB surveys, therefore, these enterprises printed their characteristics in the questionnaire form of the corresponding survey.

5) Problems related to the questionnaire.

- Questions on IVT increase enterprise burden, but these questions are less useful, because only a small number of enterprises have information on IVT. Every company defines IVT in their own way, therefore, confusion between CVT and IVT courses was very common. Respondents were confused by the following IVT condition "The main activity of person should be to study or to train, leading to a formal qualification recognized at the national level".
- We recommend adding a criterion of "time" to the defined criteria for CVT courses; otherwise enterprises sometimes mark several hours of a long lecture as a course.
- There were difficulties with defining "planned in advanced". Any other forms of training are not pre-planned, and 1-7 % (depending on the size of the company) of respondents mentioned that in enterprise there is written plan or programme for vocational training.